ABSTRACT

Title of dissertation: ANALYSIS OF REPEATED MEASURES IN THE PRESENCE OF MISSING OBSERVATIONS DUE TO DROPOUT

Jing Li, Doctor of Philosophy, 2013

Dissertation directed by: Professor Paul J. Smith
Department of Mathematics
Dr. Yi Tsong
U.S. Food and Drug Administration

Incomplete data is common in both observational studies and clinical trials. Ignoring missing data may produce seriously biased estimators and could lead to misleading results. During the last three decades, a vast amount of work has been done in this area. The approaches can be classified into the following main categories: imputation methods, likelihood-based methods and inverse probability weighting methods. Longitudinal and crossover studies with repeated measures are particularly subject to missing observations. Various methods, including generalized estimating equations (GEE) (Liang and Zeger, 1986), weighted GEE (WGEE) (Robins, Rotnitzky and Zhao, 1995) and multiple imputations, have been proposed to cope with missing data in longitudinal studies. However, very few researchers have explored the missing data issue in crossover studies. In addition to reviewing and critiquing the methods dealing with missing observations in general and in repeated measures, in this dissertation, we propose a new weighting approach for GEE

to estimate the regression parameters in crossover studies. The proposed method provides consistent and asymptotically normally distributed estimators. Simulation and asymptotic efficiency results indicate that the proposed estimators are more efficient than both regular GEE and WGEE. Applications of the proposed method are illustrated with real data.

ANALYSIS OF REPEATED MEASURES IN THE PRESENCE OF MISSING OBSERVATIONS DUE TO DROPOUT

by

Jing Li

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2013

Advisory Committee:
Dr. Paul J. Smith, Co-Advisor
Dr. Yi Tsong, Co-Advisor
Dr. Mei-Ling Ting Lee
Dr. Francis Alt
Dr. Xin He

To my wife Casey and son Aaron

Acknowledgments

I would like to express my gratitude to all those who made it possible for me to complete this dissertation.

I especially want to thank my advisors, Professor Paul Smith and Dr. Yi Tsong for their constant guidance during my doctoral study. Their perpetual energy and enthusiasm for research motivated me to go through the difficult times in my research. Without their support, this work would have been impossible.

Drs. Mei-Ling Ting Lee, Frank Alt, and Xin He deserve special thanks as my thesis committee members.

The tremendous help from the faculty members and staff of the Department of Mathematics is deeply appreciated. I would like to thank Dr. Constantina Trivisa and Ms. Alverda McCoy for their support. I enjoyed the companionship of the students from the Department of Mathematics and the statisticians at the Office of Biostatistics, FDA. I really appreciate their company and help.

I would like to give my deepest gratitude to my wife Casey, my parents and parents-in law whose patient love and unselfish support enabled me to complete this work. I am also so thankful for the joy that my son Aaron brings.

Finally I would like to thank my Lord for guiding me through this trying but very rewarding journey.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Missing data is a common problem in both observational and controlled studies. Ignoring missing data may produce seriously biased estimators and could be lead to misleading results. During the last three decades, a large amount of research has been done in analyzing incomplete data. As a result, a rich taxonomy of missing data concepts and methods as well as various data analysis tools have been developed.

One way to deal with missingness is the complete-case (CC) method where cases with any missing values are simply discarded. Advantages of this method are simplicity of implementation and the fact that valid inferences can be obtained when the data are missing completely at random (MCAR). However, when data are missing at random (MAR) but not MCAR, the CC method will produce biased estimates. As a result, several approaches have been proposed to improve CC analysis. These are likelihood-based methods (Little and Rubin, 2002), imputation methods (Rubin, 1987; Schafer, 1997), and the inverse probability weighting (IPW) method and its efficient version (Zhao and Lipsitz, 1992; Robins, Rotnitzky and Zhao, 1994; Carpenter, Kenward, Vansteelandt, 2006). In the likelihood-based method, the marginal likelihood of the observed data is maximized. In the imputation method, the incomplete cases are imputed with the conditional mean of the

missing part given the observed part. It's also possible to impute random values. The regular analysis model then can be applied. In the IPW method, the inverse of the probability that a record is missing given the observed values is used to adjust the estimating equation; thus a model for the missingness indicator R is needed, where $R = 1$ if an outcome is observed and $R = 0$ otherwise.

Longitudinal studies are characterized by a sample design which specifies repeated observations on the same experimental unit. Failure to obtain a full set of observations on a given subject results in incomplete data and/or unbalanced designs. Such missingness is a common problem in longitudinal studies. The analysis of non-Gaussian longitudinal data is difficult partly because few models for the joint distribution of the repeated observations for a subject are available. On the other hand, longitudinal data offer the advantage that data from distinct subjects are independent. In longitudinal studies, outcomes that are repeatedly measured over time may be correlated and some may be missing. Liang and Zeger (1986) proposed the generalized estimating equation (GEE) approach for longitudinal data, whose solutions are consistent. Their methods make assumptions about the mean and variance but not necessarily about the full distribution of a random variable. The estimates obtained from solving the GEE are consistent under MCAR, but might be biased under MAR. Robins, Rotnitzky and Zhao (1995) proposed a weighted generalized estimating equation (WGEE) method, which is in essence an application of the IPW method to repeated measures, for obtaining unbiased estimates in analyzing incomplete longitudinal data under MAR.

A crossover study is distinguished from a parallel group study by each subject

receiving a sequence of experimental treatments. The main advantage is that the treatments are compared within subjects. Therefore, every subject provides a direct comparison of the treatments that he or she has received, eliminating many unknown confounders. Crossover studies have been extensively used in clinical studies. As in longitudinal studies, repeated measures are taken on the same subject, and therefore missing data is also a common problem in crossover studies. However, few research results have been reported in this area.

We propose a new weighting approach for GEE to estimate the regression parameters in crossover studies. Simulations and asymptotic efficiency results indicate that the proposed estimators are more efficient than both regular GEE and WGEE. Applications of the proposed method to real data are also discussed.

This dissertation is organized as follows. In Chapter 2, we review the complete-case method and other conventional approaches to missing data analysis. Missing data methods used in repeated measures, especially for longitudinal data, are described in Chapter 3. A new weighted estimator for improvement of estimation efficiency is introduced in Chapter 4. In Chapter 5, simulation studies are performed to validate the theoretical results. Applications to real data are also shown in Chapter 5. In Chapter 6, we summarize the results and also give the potential direction of future work.

Chapter 2

Methods to Handle Missing Data

2.1 Notation and Terminology

The following terminology is based on the standard framework coined by Rubin (1976) and Little and Rubin (2002). Assume that for each independent unit in the study, we have observations $Y_{ij}, i = 1, ..., n, j = 1, ..., T$, written in vector notation as $Y_i = (Y_{i1}, ..., Y_{iT})'$, where i indicates the subject and j the measurement occasion. Define X_i as covariates, which throughout this dissertation are always observed. Generally, X_i can have missing values as well. Let R_{ij} denote missing data indicators, where $R_{ij} = 1$ if Y_{ij} is observed, and 0 otherwise. Define the vector $R_i = (R_{i1}, ..., R_{iT})'$, parallel to Y_i. Let π_{ij} denote the observation probability. Define the vector $\pi_i = (\pi_{i1}, ..., \pi_{iT})'$. One can partition Y_i into two subvectors: Y_i^o, the vector of observed values, and Y_i^m, the vector of missing values. The full data (X_i, Y_i, R_i) $i = 1, ..., n$, consist of the covariates, the complete data, together with the missing data indicators.

A nonresponse process is said to be *missing completely at random (MCAR)* if the missingness is independent of observed and unobserved data. That is,

$$P(R_{ij} = 1|X_i, Y_i) = P(R_{ij} = 1). \tag{2.1}$$

The nonresponse process is said to be *missing at random (MAR)* if the miss-

ingness is independent of the unobserved data conditional on the observed data. That is

$$P(R_{ij} = 1|Y_i^o, Y_i^m, X_i) = P(R_{ij} = 1|Y_i^o, X_i). \tag{2.2}$$

The third missingness mechanism is called *missing not at random (MNAR)* where the probability that a measurement is missing depends on unobserved data.

2.2 Complete-Case Method

A commonly used treatment of missing data in many statistical software systems is complete-case (CC) analysis, where the records with missing values are simply deleted. Then the data used in the analysis only includes the complete cases. When the density of Y, denoted by $P(Y|X;\theta)$, is known, where θ is the vector of parameters, the maximum likelihood method can be applied. The likelihood function is:

$$L_{CC}(\theta) = \prod_{i=1}^{n} P(Y_i|X_i;\theta)^{R_i}. \tag{2.3}$$

The corresponding estimating equation is:

$$U_{CC}(\theta) = \sum_{i=1}^{n} R_i U(Y_i|X_i;\theta) = 0, \tag{2.4}$$

where $U(Y_i|X_i;\theta) = (\partial/\partial\theta)\log P(Y_i|X_i;\theta)$

When the missingness mechanism is missing completely at random (MCAR), that is, π_i is independent of Y and X, then the above estimating equation is unbiased

as shown below:

$$\begin{aligned} E[U_{CC}(\theta)] &= \sum_{i=1}^{n} E[E\{R_i U_i(Y_i|X_i;\theta)|Y_i, X_i\}] \\ &= \sum_{i=1}^{n} E[U_i(Y_i|X_i;\theta)E(R_i|Y_i, X_i)] \\ &= \sum_{i=1}^{n} E[U_i(Y_i|X_i;\theta)]E(R_i|X_i) \\ &= 0. \end{aligned} \tag{2.5}$$

Thus, the estimator obtained from CC analysis, $\hat{\theta}_{CC}$, is consistent (Lehmann and Casella, 1998), and $\sqrt{n}(\hat{\theta}_{CC} - \theta_0)$ is asymptotically normally distributed with mean zero and variance:

$$\begin{aligned} \Sigma_{CC} &= E\left[R_i \frac{\partial U_i(Y_i|X_i,\theta)}{\partial\theta}\right]^{-1} E[R_i U_i(Y_i|X_i,\theta)U_i^T(Y_i|X_i,\theta)]E\left[R_i \frac{\partial U_i^T(Y_i|X_i,\theta)}{\partial\theta}\right]^{-1} \\ &= \pi_i^{-1} E[U_i(Y_i|X_i,\theta)U_i^T(Y_i|X_i,\theta)]^{-1}. \end{aligned}$$

When the missingness mechanism is not MCAR, $E(R_i|Y_i, X_i)$ will depend on Y_i. Thus equation (2.5) will not hold and the estimator $\hat{\theta}_{CC}$ will be biased (Little and Rubin, 2002). On the other hand, even if the complete-case method yields a consistent estimator when R is independent of Y, discarding incomplete cases seems to be an unnecessary waste of information.

2.3 Likelihood-based Method

Compared to the complete-case method, an obviously improved approach is the maximum likelihood (ML) method. The likelihood can be written as follows:

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n}[P(R_i=1|Y_i^o,Y_i^m,X_i)P_\theta(Y_i^o|Y_i^m,X_i)P(Y_i^m|X_i)]^{R_i} \\ &\quad \times \left[\int P(R_i=1|Y_i^o,y_i^m,X_i)P_\theta(Y_i^o|y_i^m,X_i)P(y_i^m|X_i)dy_i^m\right]^{1-R_i}. \quad (2.6) \end{aligned}$$

Under MAR, $P(R_i|Y_i^o,Y_i^m,X_i) = P(R_i|Y_i^o,X_i)$, and $P(R_i|Y_i^o,Y_i^m,X_i)$ and $P(Y_i^m|X_i)$ contain no information on θ. Thus the above joint distribution becomes

$$\begin{aligned} L(\theta) &\propto \prod_{i=1}^{n}[P_\theta(Y_i^o|Y_i^m,X_i)]^{R_i}[\int P_\theta(Y_i^o|y_i^m,X_i)P(y_i^m|X_i)dy_i^m]^{1-R_i} \\ &= \prod_{i=1}^{n}[P_\theta(Y_i^o|Y_i^m,X_i)]^{R_i}[P_\theta(Y_i^o|X_i)]^{1-R_i}. \quad (2.7) \end{aligned}$$

The MLE $\hat{\theta}_{ML}$ for θ can be obtained by solving the following estimating equation:

$$U_{ML}(\theta) = \sum_{i=1}^{n}[R_i(Y_i^o|Y_i^m,X_i;\theta) + (1-R_i)U(Y_i^o;\theta)] = 0, \quad (2.8)$$

where

$$U(Y_i^o|Y_i^m,X_i;\theta) = (\partial/\partial\theta)\log P_\theta(Y_i^o|Y_i^m,X_i),$$

and

$$U(Y_i^o|X_i;\theta) = (\partial/\partial\theta)\log P_\theta(Y_i^o|X_i).$$

This is an unbiased estimating equation. Therefore, the estimator $\hat{\theta}_{ML}$ is consistent and $\sqrt{n}(\hat{\theta}_{ML}-\theta_0)$ is asymptotically normally distributed with mean 0 and variance

$$\Sigma_{ML} = \left\{E[\pi(Y_i^o,X_i)(-\frac{\partial}{\partial\theta}U(Y_i^o|Y_i^m,X_i;\theta)) + (1-\pi(Y_i^o,X_i))(-\frac{\partial}{\partial\theta}U_\theta(Y_i^o|X_i))]\right\}^{-1}$$

If the fully parametric models can be correctly specified, the estimators are usually more efficient. Generally it's straightforward to fit a multivariate normal model using standard statistical packages. However, likelihood-based estimation relies heavily on parametric model assumptions and the ignorability assumption, which generally cannot be jointly checked from the observed data. Thus estimates are sensitive to model misspecification. In addition, even with complete data, if the focus is on models for the marginal distribution of the response, then fully parametric models for certain non-Gaussian data that preserve the marginal expectation of Y_{it} given X_i can often be cumbersome and computationally difficult when X_i is multivariate with continuous components (Prentice, 1988).

2.4 Multiple Imputation (MI) Method

Multiple imputation is a Monte Carlo approach to the analysis of incomplete data. Multiple imputation shares the same underlying philosophy as the EM algorithm and data augmentation: solving an incomplete-data problem by repeatedly estimating and solving the complete-data version (Molenberghs and Kenward, 2007). In multiple imputation, the unknown missing values of Y_m are replaced by simulated values $Y^m_{(1)}, Y^m_{(2)}, ..., Y^m_{(D)}$. Each of the D completed data sets is analyzed by standard complete-data methods. The variability among the results of the D analyses provides a measure of the uncertainty due to missing data, which, when combined with measures of ordinary sample variation, lead to a single inferential statement about the parameters of interest.

Theoretical Justifications

At the heart of the MI method is a Bayesian argument. The idea, first proposed by Rubin (1978), is to relate the observed-data posterior distribution to the complete-data posterior distribution that would have been obtained if we had observed the missing data. Then the posterior distribution for θ, the parameter of interest, is given by:

$$\begin{aligned} P(\theta|Y^o) &= \int \frac{P(\theta, Y^m, Y^o)}{P(Y^o)} dY^m \\ &= \int P(\theta, Y^m|Y^o) dY^m \\ &= \int P(\theta|Y^o, Y^m) P(Y^m|Y^o) dY^m \\ &= E_{Y^m|Y^o}[P(\theta|Y^o, Y^m)]. \end{aligned} \tag{2.9}$$

When the posterior mean and variance are adequate summaries of the posterior distribution, the content of equation (2.9) can be effectively represented by the posterior mean and variance. The former is given by:

$$\begin{aligned} E(\theta|Y^o) &= \int \theta \int P(\theta|Y^o, Y^m) P(Y^m|Y^o) dY^m d\theta \\ &= \int P(\theta|Y^o) \int \theta P(\theta|Y^o, Y^m) d\theta dY^m \\ &= E_{Y^m|Y^o}[E[\theta|Y^o, Y^m]] \\ &\approx \frac{1}{D} \sum_{i=1}^{D} \hat{\theta}_d = \hat{\theta}_{MI}, \end{aligned} \tag{2.10}$$

where D is the number of imputations and $\hat{\theta}_d$ is the maximum likelihood estimate of θ calculated by using the dth completed data set, and $\hat{\theta}_{MI}$ is the multiple imputation estimator of θ. The value of D varies, but typically researchers set $D = 5$ (Schafer, 1999).

The posterior variance is approximated by:

$$\begin{aligned}
\text{var}(\theta|Y^o) &= \int \theta^2 \int P(\theta|Y^o, Y^m)P(Y^m|Y^o)dY^m d\theta \\
&\quad - [P(\theta|Y^o, Y^m)P(Y^m|Y^o)dY^m d\theta]^2 \\
&= \int P(Y^m|Y^o) \left[\int \theta^2 P(\theta|Y^m, Y^o)d\theta - \left(\int \theta P(\theta|Y^m, Y^o)d\theta \right)^2 \right] dY^m \\
&\quad + \int P(Y^m|Y^o) \left[\int \theta P(\theta|Y^o, Y^m)d\theta \right]^2 dY^m \\
&\quad - \left[\int P(\theta|Y^m|Y^o) \left(\int \theta P(\theta|Y^o, Y^m)d\theta \right) dY^m \right]^2 \\
&= E[Var(\theta|Y^m, Y^o)|Y^o] + Var[E(\theta|Y^m, Y^o)|Y^o] \\
&\approx \frac{1}{D}\sum_{d=1}^{D} V_d + \frac{1}{D-1}\sum_{d=1}^{D}(\hat{\theta}_{MI} - \hat{\theta}_d)^2 = \bar{V} + B, \qquad (2.11)
\end{aligned}$$

where V_d is the complete-data posterior variance calculated from the dth data set, $\bar{V} = 1/D \sum_{d=1}^{D} V_d$ is the average of V_d over the imputed data sets and $B = [1/(D-1)] \sum_{d=1}^{D} (\hat{\theta}_{MI} - \hat{\theta}_d)^2$ is the between-imputation variance.

Making Proper Imputations

Multiple imputation is attractive because it can be highly efficient even for small values of D. The efficiency of an estimate based on D imputations is approximately $(1 + \gamma/D)^{-1}$, where γ is the fraction of missing information for the quantity being estimated. Therefore, there is thought to be little advantage in producing and analyzing more than a few imputed data sets (Little and Rubin, 2002).

For general missing data patterns with a multivariate normal imputation model, small-sample draws can be constructed using Markov Chain Monte Carlo (MCMC) methods (Schafer, 1997, Section 5.4). Schafer (1997), along with many software

packages including SAS/STAT, uses data augmentation, one type of MCMC, to get random draws from the posterior distribution of the parameter of interest θ. Closely related to Gibbs sampling, data augmentation is an iterative method of simulating the posterior distribution of θ. Start with an initial draw from an approximation to the posterior distribution of θ. The EM estimate of θ is usually a good starting value. Given a value of drawn at iteration t:

I-Step: Draw $Y^m_{(t+1)}$ from density $p(y^m_{(t+1)}|Y^o, \theta_{(t)})$;

P-Step: Draw $\theta^{(t+1)}$ from density $p(\theta|Y^o, Y^m_{(t+1)})$.

The iterative procedure can be shown eventually to yield a draw from the joint posterior distribution of (Y^m, θ) given Y^o in the sense that as t tends to infinity, this sequence converges to a draw from the joint distribution of (Y^m, θ) given Y^o .

Specifically, in our example the model for $p(Y^m|Y^o)$ is known as the imputation model. For the resulting estimates of θ to have the correct frequentist properties, in terms of consistency and correct coverage of confidence intervals, careful choice of $p(Y^m|Y^o)$ is crucial. Additionally, the variables that are included in Y^o must include all the variables that make the response missing at random.

Single imputation methods are simple to use, but in general they might not conform to statistical principles for making inferences, particularly when sources of uncertainty are concerned in determining treatment effect.

Multiple imputation (MI) methods allow using large amounts of information about auxiliary variables that are not included in the analysis model but can be used in the imputation model. This generally would lead to more accurate imputations. In addition, MI methods are easy to use and are available in many

standard statistical software packages, such as PROC MI and MIANALYZE in SAS. However, MI does rely on parametric assumptions, which generally cannot be verified from the observed data. Furthermore, the data model $P(Y|X,\theta)$ may be incompatible with the imputation model $P(Y|V,X;\phi)$, where V are the auxiliary variables. Compatible models have the property that the data model is what is left when auxiliary variables are integrated out of the imputation model, that is, $\int P(Y|V,X;\phi)P(V|X;\phi)dV = P(Y|X;\phi)$.

As the number of imputations (D) increases, the results from the likelihood method and multiple imputation method should converge (Molenberghs and Kenward,2007). When the number of imputations is finite, the MLE are actually more efficient although the differences may be very small.

When covariates are missing, MI offers an intuitively attractive and very manageable method for dealing with potentially very complex problems for which likelihood analyses maybe impracticable or, at least, very awkward.

As Meng (1994) shows, it is one of the great strengths of MI that the imputation and substantive models do not have to be the same (usually the imputation model is more general than the analysis model). Meng introduces the term 'uncongenial' for an imputation model which is not consistent with the substantive model. It's with these that MI has much to offer in the current setting.

2.5 Inverse Probability Weighting Method

As we know, random selection of the available data may yield biased estimators. A similar problem occurs when the observations come from a complex survey design wherein the subjects are sampled from a finite population with unequal selection probabilities. Analyzing such design-based data as a simple random sample can also introduce bias. To reduce selection bias, the inverse probability weighting (IPW) idea was introduced by Horvitz and Thompson (1952). The key idea of IPW is straightforward and intuitively attractive; that is, by weighting each subject in the data with inverse selection probability, one can produce unbiased estimators. Based on the IPW idea, Zhao and Lipsitz (1992) proposed a weighted estimating equation for general regression.

2.5.1 Inverse Probability Weighting (IPW)

The estimating equation of Zhao and Lipsitz (1992) has the following form:

$$U_W(\theta) = \sum_{i=1}^{n} \frac{R_i}{\pi_i(Y_i, X_i)} U(Y_i|X_i;\theta) = 0, \tag{2.12}$$

where $\pi_i(Y_i, X_i)$ is the observation probability and $U(Y_i|X_i;\theta)$ is the quasi-score function of $P(Y_i|X_i;\theta)$.

Zhao and Lipsitz (1992) showed that the estimator $\hat{\theta}_W$ is consistent and asymptotically normal with variance $\Sigma_W = \mathcal{I}_v^{-1}\Sigma_\theta\mathcal{I}_v^{-1}$, where $\mathcal{I}_v = E[-(\partial/\partial\theta)U(Y_i|X_i;\theta)]$, and $\Sigma_\theta = E\left[1/\pi(Y,X)U(Y|X;\theta)U(Y|X;\theta)^T\right]$.

When π_i is unknown, which is the case in many situations, it can be estimated

in conjunction with estimating θ. A parametric model (Zhao, Lipsitz and Lew 1996) can be built for π_i, for example, $\pi_i(\alpha) = f(R|Y_i^o, X_i; \alpha)$, where f is a known function indexed by unknown parameter α. Then θ and α can be estimated simultaneously by solving the following estimating equations:

$$U_W(\theta, \alpha) = \sum \frac{R_i}{\pi_i(Y_i^o, X_i; \alpha)} U(Y_i|X_i; \theta) = 0,$$

and

$$\begin{aligned} U(\alpha) &= \sum_{i=1}^{n} U(R_i|Y_i^o, X_i; \alpha) \\ &= \sum \frac{R_i - \pi_i}{\pi_i(\alpha)(1-\pi_i(\alpha))} \frac{\partial \pi_i(\alpha)}{\partial \alpha} = 0. \end{aligned} \tag{2.13}$$

Zhao, Lipsitz and Lew (1996) showed that $(\hat{\theta}_{W,\alpha}, \hat{\alpha})$ has the following asymptotic distribution

$$\sqrt{n} \begin{pmatrix} \hat{\theta}_{W_\alpha} - \theta_0 \\ \\ \hat{\alpha} - \alpha_0 \end{pmatrix} \xrightarrow{d} N\left(0, \Sigma_{\theta,\alpha}\right), \tag{2.14}$$

where $\Sigma_{\theta,\alpha}$ has the form of $\mathcal{I}^{-1}\Sigma\mathcal{I}^{-1}$,

$$\begin{aligned} \mathcal{I} &= \begin{pmatrix} (\partial/\partial\theta)U_i^W(Y|X;\theta,\alpha) & (\partial/\partial\alpha)U_i^W(Y|X;\theta,\alpha) \\ \\ (\partial/\partial\theta)U_i(R|Y_i^o, X_i;\alpha) & (\partial/\partial\alpha)U_i(R|Y_i^o, X_i;\alpha) \end{pmatrix} \\ &= \begin{pmatrix} \mathcal{I}_v & \Omega_{12} \\ \\ 0 & \mathcal{I}_\alpha \end{pmatrix} = \mathcal{I}(\theta, \alpha), \end{aligned}$$

and

$$\Sigma = \text{var}\left(\frac{1}{\sqrt{n}}(U_W(\theta,\alpha), U(\alpha))^T\right) = \begin{pmatrix} \Sigma_\theta & \Omega_{12} \\ \\ \Omega_{12}^T & \mathcal{I}_\alpha \end{pmatrix}.$$

Thus we have

$$\sqrt{n} \begin{pmatrix} \hat{\theta}_{W_\alpha} - \theta_0 \\ \\ \hat{\alpha} - \alpha_0 \end{pmatrix} \xrightarrow{d} N\left[\mathbf{0}, \begin{pmatrix} \mathcal{I}_v & \Omega_{12} \\ \\ 0 & \mathcal{I}_\alpha \end{pmatrix}^{-1} \begin{pmatrix} \Sigma_\theta & \Omega_{12} \\ \\ \Omega_{12}^T & \mathcal{I}_\alpha \end{pmatrix} \begin{pmatrix} \mathcal{I}_v & \Omega_{12} \\ \\ 0 & \mathcal{I}_\alpha \end{pmatrix}^{-1}\right],$$

and

$$\sqrt{n}(\hat{\theta}_{W_\alpha} - \theta_0) \xrightarrow{d} N[0, \mathcal{I}_v^{-1}(\Sigma_\theta - \Omega_{12}\mathcal{I}_\alpha^{-1}\Omega_{12})\mathcal{I}_v^{-1}]. \tag{2.15}$$

Hence, $\hat{\theta}_{W_\alpha}$ is more efficient than $\hat{\theta}_W$, which assumes the observation distribution is known.

Alternatively, a nonparametric model of π_i can be constructed. Wang et al (1997) considered nonparametric kernel smoothers for the selection probability π_i. The nonparametric estimate avoids the nonrobustness of parametric methods mentioned above. However, the nonparametric method encounters another problem of bandwidth selection. The selection of a good value of bandwidth is crucial to ensure the asymptotic unbiasedness of the estimation.

2.5.2 Improved IPW (Augmented IPW)

The IPW method does not use all observed information. Therefore $\hat{\theta}_W$ is not efficient but can be improved by incorporating all available information in incomplete cases. Robins, Rotnizky and Zhao (1994) showed that the efficiency of IPW can be improved by subtracting the projection of the estimating function onto the nuisance tangent space, which is the closed span of nuisance scores. Lipsitz, Ibrahim and Zhao (1999) further explored the properties of this approach.

The resulting estimating function has the following form:

$$U_{AW} = \sum_i \left[\frac{R_i}{\pi_i} U_{\theta,i}(Y|X) - \left(\frac{R_i}{\pi_i} - 1 \right) E\left[U_{\theta,i}(Y|X) | Y_i^o, X_i \right] \right]. \tag{2.16}$$

Since $E(U_{AW}) = 0$, the estimate $\hat{\theta}_{AW}$ is consistent. Using the followings

$$-\frac{1}{n}\frac{\partial}{\partial\theta} U_{AW}(\theta) \xrightarrow{p} \mathcal{I}_v,$$

$$\begin{aligned}\text{var}(\frac{1}{\sqrt{n}}U_{AW}(\theta)) &= \mathbf{E}\left[\frac{1}{\pi}U_\theta(Y|X)U_\theta(Y|X)^T\right] - \mathbf{E}\left[\frac{1-\pi}{\pi}U_\theta(Y|X)U_\theta(Y|X)^T\right] \\ &= \Sigma_\theta - \mathbf{E}\left[\frac{1-\pi}{\pi}U_\theta(Y|X)U_\theta(Y|X)^T\right] = \Sigma_{AW}.\end{aligned}$$

Therefore, $\sqrt{n}(\hat{\theta}_{AW} - \theta_0)$ is asymptotic normally distributed with mean 0 and variance $\mathcal{I}_v^{-1}\Sigma_{AW}\mathcal{I}_v^{-1}$. It is clear that the improved inverse probability weighting method reduces the variance of the IPW method.

Often, $P(Y^m|Y^o, X)$ and $P(R = 1|Y^o, X)$ are unknown. As shown in previous sections dealing with the imputation methods, $P(Y^m|Y^o, X)$, the imputation model, needs to be correctly specified. In the inverse probability weighting method, $P(R = 1|Y^o, X)$, the model for the probability of observing the data, needs to be correct. However, in the augmented IPW method, a consistent estimator of θ can be obtained if either, not necessarily both, of $P(R = 1|Y^o, X)$ or $P(Y^m|Y^o, X)$ is correctly specified. This property is called double robustness. Therefore, these estimators enjoy greater robustness against model misspecification than both imputation and IPW estimators.

Specifically, efficient IPW estimators require three models: (1) The substantive model which relates the outcome to explanatory variables/covariates of interest. (2) A model for the probability of observing the data. This is usually a logistic model of some form. (3) A model for the joint distribution of the partially and fully observed data, which is compatible with the substantive model in (1).

The interesting property of (2.16) is that if either model (2) or (3) is misspecified, but not both, the estimators in model (1) remain consistent. By contrast, if model (2) is wrong, the inverse probability weighting estimator will be inconsistent,

whereas multiple imputation will generate inconsistent estimators if model (3) is not correctly specified. Lipsitz et al. (1999) proposed parametric approaches to specify those distributions under the assumption of MAR. The steps to improve the efficiency are also discussed in Rotnitzky and Robins (1997).

Inverse probability weighting is generally simple to implement and available in standard statistical packages with weighted analyses. It's in essence a semiparametric method, making assumptions only about the mean and variance. Thus it tends to be more robust than the likelihood-based method, which requires assuming a parametric model for the full-data response distribution. An important advantage of the IPW method is that under a correctly specified model for observation, many auxiliary variables can be accommodated, including information on previously observed outcomes. This would generally improve the prediction of future outcomes being observed or missed. It's able to handle reasonably well the missing data in discrete variables. In addition, IPW method can be extended to estimating quantities other than the mean, such as the median.

Generally the IPW estimator is not efficient. But its efficiency can be significantly improved by AIPW, which also possesses the doubly-robust property. Another potential disadvantage of the IPW method is that it needs to make assumptions about the observation probability. Usually a logistic regression model is assumed. However, the possibility of correctly specifying a probability model for observations is much greater than specifying a correct response model.

Chapter 3

Missing Data Analysis in Longitudinal Study

3.1 Notation and Assumptions

Consider a repeated measure study where subjects are assessed over a fixed interval from time 1 to T. Let $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{iT})^T$ be the vector of outcome variables corresponding to subject i, $i = 1, 2, \ldots, n$, measured at each time t. Let $X_i = (X_{i1}^T, X_{i2}^T, \ldots, X_{iT}^T)^T$, where X_{it} is a vector of explanatory variables associated with Y_{it} and includes the constant 1 as a component. Here X_{it} may be a vector of baseline explanatory variables X_i^*, such as, treatment indicator, gender, age, and possibly pretreatment clinical status, or a deterministic function of time and the baseline variables, e.g., $X_{it} = X_i^* t$. It's assumed that X_i is completely observed for every subject. The marginal distribution of Y_{it} given X_i is

$$E(Y_{it}|X_i) = g^{-1}(\eta_{it}) = g^{-1}(X_i^T\theta), \tag{3.1}$$

where θ is a vector of unknown parameters and g is a known link function.

Due to various reasons, the full vector Y_i is not always observed. Define $R_i = (R_{i1}, R_{i2}, ..., R_{iT})^T$ to be the observation indicator for Y_i. Let $R_{it} = 1$ if subject i is observed at time t and $R_{it} = 0$ otherwise. The missing data pattern is assumed to be monotone, that is, Y_{ik} is not observed if Y_{ij} is missing, where $k > j$, which is often the case in repeated measure studies, such as longitudinal studies and

crossover studies. Let $\bar{W}_{it} = \{X_i, Y_{i1}, \ldots, Y_{i(t-1)}\}$ be the entire history of subject i before time t.

A nonresponse process is said to be *missing completely at random (MCAR)* if the missingness is independent of both observed and unobserved data. That is,

$$P(R_{it} = 1 | R_{i(t-1)} = 1, \bar{W}_{it}, Y_{it}) = P(R_{it} = 1 | R_{i(t-1)} = 1). \tag{3.2}$$

The nonresponse process is said to be *missing at random (MAR)* if the missingness is independent of the unobserved data conditional on the observed data, that is

$$P(R_{it} = 1 | R_{i(t-1)} = 1, \bar{W}_{it}, Y_{it}) = P(R_{it} = 1 | R_{i(t-1)} = 1, \bar{W}_{it}). \tag{3.3}$$

Another missingness mechanism is *missing not at random (MNAR)* where the probability of a measurement being missing depends on unobserved data.

Throughout, we shall assume MAR without stating otherwise and that the observation probability is bounded away from zero:

$$P(R_{it} = 1 | R_{i(t-1)} = 1, \bar{W}_{it}) > \sigma > 0 \tag{3.4}$$

3.2 Last Observation Carried Forward (LOCF)

LOCF has been a commonly used single imputation method in longitudinal study for the last few decades. It's based on the strong assumption that the outcome of a participant does not change after dropout, which is generally not the case in reality. Although using LOCF relies on the plausibility of the assumptions underpinning these estimators, the pragmatic justification often stems from the sometimes

mistaken view that it provides a simple and conservative imputation that will help prevent approval of ineffective treatments (National Research Council, 2010). However this is not necessarily the case, since, for example, LOCF is anticonservative in situations where participants off study treatment generally do worse over time. In such cases, if many participants discontinue study treatment due to problems with tolerability, the treatment can be made to look much better than the control by such an imputation strategy.

3.3 Generalized Estimating Equations (GEE)

The analysis of non-Gaussian longitudinal data is difficult partly because few models for the joint distribution of the repeated observations for a subject are available. On the other hand, longitudinal data offers the advantage that data from distinct subjects are independent. In longitudinal studies, outcomes that are repeatedly measured over time may be correlated and some may be missing. Liang and Zeger (1986) proposed the generalized estimating equation (GEE) approach to longitudinal data, whose solutions are consistent for θ provided only that the model for the marginal means of the outcomes at each occasion is correctly specified. This approach is an extension of quasi-likelihood methods (McCullagh and Nelder, 1989) to the multivariate regression setting and results in reweighted least squares estimators of the parameter θ. Assuming that the marginal mean μ_i has been correctly specified as $g(\mu_i) = \eta_i = X_i^T\theta$, where g is a known link function, the estimating

equation for θ in the presence of missing data is

$$U(\theta, \alpha) = \sum_{i=1}^{n} \frac{\partial \mu_i^T}{\partial \theta} V_i(\theta, \alpha)^{-1} C_i (Y_i - \mu_i) = 0, \tag{3.5}$$

where $V_i(\theta, \alpha) = \phi A_i(\mu_i)^{1/2} \Omega_i(\rho) A_i \mu_i{}^{1/2}$, $A_i = \text{diag}\{\text{var}(y_{i1}), \text{var}(y_{i2}), ..., \text{var}(y_{iT})\}$, $\alpha = (\phi, \rho)$, $\Omega_i(\rho)$ is a "working" correlation matrix of Y_i, and

$C_i = \text{diag}\{R_{i1}, R_{i2}, ..., R_{iT}\}$.

One can choose $\Omega_i(\rho)$ as the identity matrix or the equicorrelation matrix and the specified correlation structure need not be the correct correlation structure of Y_i. The GEE estimator, $\hat{\theta}_G$, is the root of $U(\theta, \hat{\alpha}) = 0$, where $\hat{\alpha}$ satisfies $\sqrt{n}(\hat{\alpha} - \alpha^*) = O_p(1)$. $\hat{\theta}_G$ is asymptotically normally distributed with mean θ and variance:

$$\Sigma_G = \left(\sum_{i=1}^{n} D_i^T V_i^{-1} D_i \right)^{-1} \left\{ \sum_{i=1}^{n} D_i^T V_i^{-1} \text{var}(Y_i) V_i^{-1} D_i \right\} \left(\sum_{i=1}^{n} D_i^T V_i^{-1} D_i \right)^{-1}, \tag{3.6}$$

where$D_i = \partial \mu_i^T / \partial \theta$.

Note that the consistency of $\hat{\theta}_G$ and Σ_G depends only on the correct specification of the mean, not on the correct choice of the "working" correlation matrix. However the correct specification of the "working" correlation matrix will help improve the efficiency. In addition, when the "working" correlation matrix represents the true correlation, the assumption of MCAR can be unnecessary. In summary, Liang and Zeger (1986) proposed the generalized estimating equation (GEE) approach to longitudinal data, whose solutions are consistent under MCAR. The methods avoids the need for multivariate distributions by only assuming a functional form of the marginal distribution at each occasion.

3.4 Weighted Generalized Estimating Equations (WGEE)

Liang and Zeger (1986) pointed out that $\hat{\theta}_G$, the solution by solving the GEE, is consistent under MCAR but not under MAR. Robins, Rotnitzky and Zhao (1995) proposed a weighted generalized estimating equation (WGEE) for obtaining unbiased GEE estimates under MAR.

Let $\bar{\lambda}_{it} = P(R_{it} = 1|R_{i(t-1)}, \bar{W}_{it})$ be the probability of observing a response at time t. Typically, $\bar{\lambda}_{it}(\alpha)$ could be a logistic function. Let $\hat{\alpha}$ be the partial maximum likelihood estimator (MLE) that maximizes the partial likelihood estimator,

$$L(\alpha) = \prod_i L_i(\alpha) = \prod_i \prod_t [\bar{\lambda}_{it}(\alpha)^{R_{it}} \{1 - \bar{\lambda}_{it}(\alpha)\}^{1-R_{it}}]^{R_{i(t-1)}}. \tag{3.7}$$

The contribution to the score for α from the ith subject is

$$S_i(\alpha) = \left\{ \frac{\partial}{\partial \alpha} L_i(\alpha) \right\} = \sum_{t=1}^{T} (R_{it} - \bar{\lambda}_{it}(\alpha) R_{i(t-1)}) \frac{\partial}{\partial \alpha} \text{logit} \bar{\lambda}_{it}\{(\alpha)\}. \tag{3.8}$$

Let $\Sigma = \text{var}\{S_i(\alpha)\}$ be the asymptotic variance of $\sqrt{n}(\hat{\alpha} - \alpha)$. Then $S_i(\alpha)$ simplifies to $\sum_{t=1}^{T}(R_{it} - \bar{\lambda}_{it}(\alpha)R_{i(t-1)})h(\bar{W}_{it})$ if $\bar{\lambda}_{it}$ follows the logistic regression model $\text{logit}(\bar{\lambda}_{it}) = \alpha^T h(\bar{W}_{it})$ for some vector function $h(\cdot)$. Define $\bar{\pi}_{it}(\alpha) = \bar{\lambda}_{i1}(\alpha) \times \cdots \times \bar{\lambda}_{it}(\alpha)$, where $\bar{\pi}_{it}(\alpha)$ is the conditional probability of observing subject i at the tth occasion given the entire vector $\bar{W}_{i(T+1)}$. Define

$$\triangle_i(\alpha) = \begin{pmatrix} R_{i1}\bar{\pi}_{i1}(\alpha)^{-1} & 0 & \cdots & 0 \\ 0 & R_{i2}\bar{\pi}_{i2}(\alpha)^{-1} & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & R_{iT}\bar{\pi}_{iT}(\alpha)^{-1} \end{pmatrix} \tag{3.9}$$

as the $T \times T$ diagonal matrix with diagonal elements $\triangle_{it}(\alpha) = R_{it}\bar{\pi}_{it}(\alpha)^{-1}$ Then the WGEE has the following form:

$$U(\theta, \hat{\alpha}) = n^{-1/2} \sum_{i=1}^{n} D_i V_i^{-1} \triangle_i(\hat{\alpha})(Y_i - \mu_i) = 0, \tag{3.10}$$

where $D_i = \partial \mu_i^T / \partial \theta, V_i$ is the "working" covariance matrix. Robins, Rotnitzky and Zhao (1995) showed that the solution to estimating equation (3.10), $\hat{\theta}_{WG}$, is consistent and asymptotically normal.

Chapter 4

Proposed Reweighted Generalized Estimating Equations Method

4.1 Motivation

While the WGEE method offers a very useful alternative to consistently estimating the repeated measures with missing observations, it does have serious disadvantages. When the observation probability $\bar{\pi}_{it}$ in WGEE is small, WGEE estimates become very unstable. Additionally, the efficiency of a WGEE estimate is sacrificed for the robustness by assuming only the mean and variance of a distribution. In order to improve upon the WGEE method, especially the efficiency of WGEE, motivated by the methods of handling missing data in survey research (Brick and Kalton, 1996), we developed a new weighted estimating equation, a reweighted approach, for repeated measures in the presence of missing observations. The proposed method handles missing observations under the MAR assumption. The basic idea of the reweighted method is first to weight each subject by the inverse of the observation probability to correct the selection bias created by missingness and then to impose a simpler selection probability on each subject. Here we suggest setting the imposed simpler selection probability as a marginal probability at each time point for longitudinal study or each period for crossover study. The selection probability is modeled parametrically. Generally it can be a logistic regression model that regresses only on the covariates of the subject. We show that the reweighted GEE (RGEE)

estimates are asymptotically consistent and normal and that RGEE estimates are asymptotically more efficient than WGEE estimates.

4.2 Reweighted GEE Estimators

Denote the newly derived weight as $q_{it} = \omega_{it}/\bar{\pi}_{it}$, where $\bar{\pi}_{it} = \bar{\lambda}_{i1} \times \ldots \times \bar{\lambda}_{iT}$, $\bar{\lambda}_{it}$ is the conditional probability of observing subject i at the tth occasion given the entire history of subject i, and ω_{it} is an observation probability for subject i at time point t by a logistic regression model that only regresses on the covariates. That is, $\text{logit}(\omega_i) = \alpha_0 + \alpha^T X_i$, where X_i are the covariates.

Hence, the reweighting estimating equation is

$$U(\theta) = n^{-1/2} \sum_{i=1}^{n} U_i(\theta) = n^{-1/2} \sum_{i=1}^{n} D_i V_i^{-1} \triangle_i^* (Y_i - \mu_i) = 0, \tag{4.1}$$

where $D_i = \partial \mu_i^T / \partial \theta$, V_i is the "working" covariance matrix, and $\triangle_i^* = \text{diag}\{1, R_{i2} \times q_{i2}, \cdots, R_{iT} \times q_{iT}\}$.

Let $Z_i^*(\psi) = (U_i^T(\theta, \alpha), S_i^T(\alpha))^T$, where $\psi^T = (\theta^T, \alpha^T)$, $U_i^T(\theta, \alpha)$ is the contribution to the score for θ, α from the ith subject, and $S_i^T(\alpha)$ is the contribution to the score for α from the ith subject. The following regularity conditions are needed for Theorems 1 and 2:

Regularity Conditions

1. θ_0 and α_0 lie in the interior of compact sets of θ and α.

2. $(\bar{R}_{i(T+1)}^T, \bar{W}_{i(T+1)}^T)^T$, $i = 1, \ldots, n$, are i.i.d.

3. $\bar{\lambda}_{it}(\alpha) > c > 0$ for all $\alpha \in \mathcal{A}, t = (1, \ldots, T)$ for some c.

4. $E_{\psi_0}[Z_i^*(\psi)] \neq 0$ if $\psi \neq \psi_0$.

5. $\text{var}[Z_i^*(\psi_0)]$ is finite and positive definite.

6. $E[(\partial/\partial\psi_T)Z_i^*(\psi)]$ exists and is invertible.

7. $E[\sup_{\psi\in\psi}\|Z_i^*(\psi)\|], E[\sup_{\psi\in\psi}\|(\partial/\partial\psi_T)Z_i^*(\psi)\|], E[\sup_{\psi\in\psi}\|Z_i^*(\psi)Z_i^*(\psi)^T\|]$ are all finite where $\|A\| \equiv \{\Sigma_{ij}A_{ij}^2\}^{1/2}$ for any matrix A with elements A_{ij} and ψ is the Cartesian product of α and θ.

8. $L_i(\psi)$ is a parametric model for the observed data, where $L_i(\psi)$ is a density that differs from the true density only in that ψ replaces ψ_0.

9. For all ψ^* in a neighborhood N of ψ_0, $E_{\psi^*}[Z_i^*(\psi)]$ and $E_{\psi^*}[\sup_{\psi\in\psi}\|Z_i^*(\psi)Z_i^*(\psi)^T\|]$ are bounded, where E_{ψ^*} refers to expectation with respect to $L_i(\psi^*)$

Theorem 1. *Under regularity conditions (1-9) and assumptions (3.3), (3.4), the solution to the reweighted estimating equations (4.1), $\hat{\theta}_{RG}$, is consistent and asymptotically normal for estimating θ. That is,*

$$\sqrt{n}(\hat{\theta}_{RG} - \theta_0) \xrightarrow{d} N(0, \Gamma^{-1}\Sigma_\theta\Gamma^{-1}), \tag{4.2}$$

where $\Gamma = E\{(\partial U_i(\theta)/\partial\theta^T)\}$ and $\Sigma_\theta = E[1/\pi(Y,X)U(Y|X;\theta)U(Y|X;\theta)^T]$.

When $\bar{\pi}_{it}$ is unknown, which is often the case in practice, it can be estimated in conjunction with estimating θ. Let the new weight be $q_{it}(\alpha) = \omega_{it}/\bar{\pi}_{it}(\alpha)$, where $\bar{\pi}_{it}(\alpha) = \bar{\lambda}_{i1}(\alpha)\times\cdots\times\bar{\lambda}_{iT}(\alpha)$, and $\bar{\lambda}_{it}(\alpha)$ is the corresponding conditional probability.

Then the reweighting estimating equations are

$$U(\theta,\alpha)=n^{-1/2}\sum_{i=1}^{n}U_i(\theta,\alpha)=n^{-1/2}\sum_{i=1}^{n}D_iV_i^{-1}\triangle_i^*(\alpha)(Y_i-\mu_i)=0, \tag{4.3}$$

and

$$S(\alpha)=\sum_{i=1}^{n}S_i(\alpha)=\sum_{i=1}^{n}\sum_{t=1}^{T}(R_{it}-\bar{\lambda}_{it}(\alpha)R_{i(t-1)})\frac{\partial}{\partial\alpha}\text{logit}\bar{\lambda}_{it}\{(\alpha)\}, \tag{4.4}$$

where $\triangle_i^*(\alpha)=\text{diag}\{1,R_{i2}\times q_{i2}(\alpha),\cdots,R_{iT}\times q_{iT}(\alpha)\}$ and $S_i(\alpha)$ is the contribution to the score for α from the ith subject.

Theorem 2. *The solution to the reweighted estimating equations (4.3), $\hat{\theta}_{RG_\alpha}$, under regularity conditions (1-9) and assumptions (3.3), (3.4), is consistent and asymptotically normal for estimating θ. That is,*

$$\sqrt{n}(\hat{\theta}_{RG_\alpha}-\theta_0)\xrightarrow{d}N(0,\Gamma^{-1}(\Sigma_\theta-B\Omega B^T)\Gamma^{-1}), \tag{4.5}$$

where $\Gamma=E\{\partial U_i(\theta,\alpha)/\partial\theta^T\},\Sigma_\theta=E\{U_i(\theta,\alpha)U_i(\theta,\alpha)^T\},B=E\{\partial U_i(\theta,\alpha)/\partial\alpha^T\}=E\{U_i(\theta,\alpha)S_i(\alpha)^T\}$, and $\Omega=E\{S_i(\alpha)S_i(\alpha)^T\}$. Moreover $\Gamma,\Sigma_\theta,B,\Omega$ can be estimated by

$$\hat{\Gamma}=(1/n)\sum(\partial\mu_i(\hat{\theta})^T/\partial\theta)V_i(\hat{\theta})\triangle_i(\hat{\alpha})(\partial\mu_i(\hat{\theta})^T/\partial\theta),$$

$$\hat{\Sigma}_\theta=(1/n)\sum U_i(\hat{\theta},\hat{\alpha})U_i(\hat{\theta},\hat{\alpha})^T,$$

$$\hat{B}=(1/n)\sum U_i(\hat{\theta},\hat{\alpha})S_i(\hat{\alpha})^T,$$

and

$$\hat{\Omega}=(1/n)\sum S_i(\hat{\alpha})S_i(\hat{\alpha})^T,$$

respectively.

4.3 Proofs

Proof of Theorem 1

When $\bar{\pi}_{it}$ is known, the reweighted estimator, $\hat{\theta}_{RG}$, the estimator obtained by solving estimating equation (4.1), is consistent (Lehmann and Casella, 1998) because

$$\begin{aligned} E[U_i(\theta)] &= E[E\{U_i(\theta)|Y_i, X_i\}] \\ &= E\left[\frac{\partial \mu_i^T}{\partial \theta} V_i^{-1} E(\triangle_i^*|Y_i, X_i)(Y_i - \mu_i)\right] \\ &= E\left[\frac{\partial \mu_i^T}{\partial \theta} V_i^{-1} W_i (Y_i - \mu_i)\right] \\ &= 0, \end{aligned}$$

where $W_i = \text{diag}\{1, \omega_{i1}, \ldots, \omega_{iT}\}$. This follows because given covariate value X_i, $\bar{\pi}_{it}$ doesn't depend on the outcome Y_i.

Under regularity conditions $(1-9)$, Theorem (3.4) of Newey and McFadden (1994) implies that with probability approaching 1, $\hat{\theta}_{RG}$ exists, is unique and satisfies

$$\sqrt{n}(\hat{\theta}_{RG} - \theta_0) = \left[-\frac{1}{n}\frac{\partial}{\partial \theta^T} U(\theta_0)\right]^{-1} \frac{1}{\sqrt{n}} U(\theta_0) + o_p(1).$$

Using large sample theory, under regularity conditions $(1-9)$, it follows that

$$-\frac{1}{n}\frac{\partial}{\partial \theta} U(\theta_0) \xrightarrow{p} E\left[-\frac{\partial}{\partial \theta} U(Y_1|X_1; \theta_0)\right] = \Gamma,$$

and

$$\frac{1}{\sqrt{n}} U(\theta_0) \xrightarrow{d} N(0, \Sigma_{\theta_0}),$$

where

$$\Sigma_{\theta_0} = E\left[\frac{1}{\pi(Y, X)} U(Y|X; \theta_0) U(Y|X; \theta_0)^T\right].$$

Thus we have

$$\sqrt{n}(\hat{\theta}_{RG} - \theta_0) \xrightarrow{d} N(0, \Sigma_{RG}),$$

where $\Sigma_{RG} = \Gamma^{-1}\Sigma_{\theta_0}\Gamma^{-1}$.

Proof of Theorem 2

As in Robins, Rotnizky and Zhao (1995) and in Newey and McFadden (1994), under regularity conditions $(1-9)$ and by applying Taylor expansion, we have

$$\sqrt{n}(\hat{\alpha} - \alpha_0) = \left\{ -E\frac{\partial S_i(\alpha)}{\partial \alpha} \right\}^{-1} S_i(\alpha_0) + o_p(1),$$

and

$$\sqrt{n}(\hat{\theta}_{RG_\alpha} - \theta_0) = \left\{ -E\frac{\partial U_i(\theta_0, \alpha_0)}{\partial \theta} \right\}^{-1} [U(\theta, \alpha) + \left\{ E\frac{\partial U_i(\theta, \alpha)}{\partial \alpha} \right\}^{-1} \sqrt{n}(\hat{\alpha} - \alpha_0)] + o_p(1).$$

As $E\{\partial U_i(\theta, \alpha)/\partial \alpha\} = -E\{U_i(\theta, \alpha)S_i(\alpha)^T\}$, which is the "Generalized Information Equality" (Pierce, 1982), and

$$\text{var}\{S_i(\alpha)\} = -E\{\partial S_i(\alpha)\partial \alpha\} = E\{S_i(\alpha)S_i(\alpha)^T\},$$

we then obtain

$$\sqrt{n}(\hat{\theta}_{RG_\alpha} - \theta_0) = \Gamma^{-1} \sum_{i=1}^{n} \text{Resid}\{U_i(\theta, \alpha), S_i(\alpha)\} + o_p(1),$$

where $\Gamma^{-1} = -E\{\partial U_i(\theta, \alpha)/\partial \alpha\}$, and

$$\text{Resid}\{U_i(\theta, \alpha), S_i(\alpha)\} = U_i(\theta, \alpha) - E\{U_i(\theta, \alpha)S_i(\alpha)^T\}[E\{S_i(\alpha)S_i(\alpha)^T\}]^{-1}S_i(\alpha)$$

is the residual from the population least squares regression of $U_i(\theta, \alpha)$ on $S_i(\alpha)$.

Therefore,

$$\text{var}\left\{\text{Resid}[U_i(\theta,\alpha)S_i(\alpha)]\right\} = \Sigma_{\theta_0} - B\Omega^{-1}B^T,$$

where $B = E\{U_i(\theta,\alpha)S_i(\alpha)^T\}, \Omega = E\{S_i(\alpha)S_i(\alpha)^T\}$. From the central limit theorem, the asymptotic distribution of $\sqrt{n}(\hat{\theta}_{RG_\alpha} - \theta_0)$ is normal with mean 0 and variance $\Gamma^{-1}(\Sigma_{\theta_0} - B\Omega^{-1}B^T)\Gamma^{-1}$.

If some components of X_i are grouping indicators, the proofs in this section can be extended to this case, provided that n_g, the number of subjects in group g, goes to infinity and $n_g \to C_g \in (0,1)$ for $g = 1,\ldots,G$.

Chapter 5

Application of the Proposed Method in Crossover Study Setting

5.1 Crossover Study and Designs

In a parallel group study, each experimental unit is randomized to receive one experimental treatment. A crossover study is distinguished from a parallel group study by each subject's receiving a sequence of experimental treatments. Note that typically the aim is still to compare the effects of individual treatments, not the sequences themselves. The main advantage is that the treatments are compared "within-subject". Therefore, every subject provides a direct comparison of the treatments that he or she has received, eliminating many unknown confounders. Crossover designs remove from the treatment (and period) comparisons any quantity that is related to the differences between the subjects as it is well known that the variability of measurements taken on different subjects is far greater than the variability of repeated measurements taken on the same subject (Jones and Kenward, 2003). Crossover studies have been extensively used in clinical studies, particularly in the chronic diseases.

Although the use of repeated measurements on the same subject provides great advantages, it also brings one potential disadvantage, which is the possibility that the effect of a treatment given in one period might still be present at the start of the following period. This phenomenon, called "carryover effect", can be reduced by

Table 5.1: Orthogonal Latin Square Design for Four Treatments

	Period			
Sequence	1	2	3	4
1	A	B	C	D
2	B	A	D	C
3	C	D	A	B
4	D	C	B	A
5	A	D	B	C
6	B	C	A	D
7	C	B	D	A
8	D	A	C	B
9	A	C	D	B
10	B	D	C	A
11	C	A	B	D
12	D	B	A	C

wash-out, in which the active effects of a treatment given in the previous period to be washed out of the body before each subject begins the next period of treatment, and by the choice of designs and combinations of them, among other methods.

In the presence of carryover effects, in order to have the highest possible efficiency the design must be balanced. The term "balance" refers to the combinatorial properties that the design must possess: (a) In a balanced design, not only does each treatment occur once with each subject, but also (b), over the whole design each treatment occurs the same number of times in each period, and (c) the number of subjects who receive treatment i in one period followed by treatment j in the next period is the same for all $i \neq j$. This can be achieved by using a complete set of orthogonal Latin squares.

A complete set of $t \times t$ orthogonal Latin squares contains $t - 1$ squares and

Table 5.2: Williams Designs for Four and Three Treatments

	Period			
Sequence	1	2	3	4
4×4 Williams Design				
1	A	B	D	C
2	B	C	A	D
3	C	D	B	A
4	D	A	C	B
3×3 Williams Design				
1	A	B	C	
2	B	C	A	
3	C	A	B	
4	A	C	B	
5	B	A	C	
6	C	B	A	

complete sets exist for values of t that are prime or are powers of a prime. A notable exception is therefore $t = 6$. Although using orthogonal Latin squares has additional advantages, they require more subjects and is generally more difficult to execute in clinical trials. Also, the loss of subjects from the complete set is likely to be more damaging as its combinatorial structure is more complex.

It was shown that balance could be achieved by a Williams design (Williams, 1949) using only one particular Latin square if t is even and by using only two particular squares if t is odd. In a Williams design, each treatment precedes every other treatment, excluding itself, equally often and thus a balance for carryover effects is achieved. For more than three treatments, a Williams design, therefore,

requires fewer subjects than complete sets of orthogonal squares. It's a design commonly used in crossover clinical trials.

As in longitudinal studies, repeated measures are taken on the same subject in crossover studies. Similarly, missing data is also a common problem. However, few studies have been reported in this area. We would like to compare the incomplete data analysis of both simulated data and industry submitted crossover clinical trial data using various approaches including full cohort, complete-case analysis, WGEE and RGEE, our proposed method. Full cohort are the full data without missing observations while the complete-case method only analyzes the subjects with observations in every period.

Let p and t denote the number of periods and treatments, respectively. Here we focus on the crossover designs with $p = t$.

5.2 Simulations

Simulations were conducted for a 2×2 crossover study and the Williams design for a 3×3 crossover study to evaluate the performances of the proposed reweighted GEE (RGEE) along with full cohort, complete-case and WGEE approaches. The primary model of interest is the marginal mean of Y_{it} conditional on X_{it},

$$E(Y_{it}|X_{it}) = \mu_{it} = g^{-1}(\eta) = g^{-1}(X_{it}^T\theta), \tag{5.1}$$

where θ is a vector of unknown parameters, g is a known link function, and X_i are covariates used in the simulations. In this chapter, the response variable is assumed

Table 5.3: Simulation Settings for 2×2 Crossover Design

Missing Prop.	Sample Size	Treatment Effect	Table
$20 - 30\%$	1000	Large *vs* Small	5.4
$40 - 50\%$	1000	Large *vs* Small	5.5
$20 - 30\%$	500 *vs* 1000	Large	5.6
$10 - 15\%$	1000	Large *vs* Small	B.4
$60 - 70\%$	1000	Large *vs* Small	B.5

to follow a normal distribution so that g is the identity function.

2×2 Crossover Study

For 2×2 crossover simulation study, we generated 500 replicated study data sets with sample sizes of $N = 500$ or 1000. The covariates are treatment, period, and baseline. It's assumed that the outcome in period 1 of each subject is always observed, while the response at period 2 outcome is observed at random according to a logistic regression model

$$\text{logit}\left\{P(R_{i2} = 1|X_i, Y_{i1})\right\} = \alpha_0 + \alpha_x^T X_i + \alpha_y Y_{i1}, \tag{5.2}$$

where X_i is the baseline covariates of subject i in period 2 and Y_{i1} is the observation in period 1.

Asymptotic variances of the estimators were estimated using the "sandwich" estimators, and the efficiency of the RGEE estimators was evaluated relative to other approaches, particularly relative to those estimated by the WGEE method. Also studied were the effects of sample size and choice of "working" covariance on the performances of each approach. Specifically, for studying the effect of sample size, the performances of the estimators were evaluated for sample sizes of 500 or 1000.

The influence of missing proportions is explored for the 2×2 crossover study. The following proportions of missing data were generated under the missing at random mechanism: $10 - 15\%$, $20 - 30\%$, $40 - 50\%$, and $60 - 70\%$. It's generally known that a large treatment effect is easily detected. To study how the proposed method performs with small treatment effects, simulations were also run on drug with small treatment effect coefficient while the other factors were held constant.

As shown in the following Q-Q plots and histograms, the estimates produced by all the methods generally follow normal distributions. All approaches except CC analysis produce consistent estimates. Not surprisingly, the full cohort analysis yields the best estimates in terms of consistency and efficiency. The RGEE estimates are the best among those estimated by all the other approaches. The bias of the CC analysis becomes more problematic as the missing proportion increases. It's also shown that the "sandwich" estimators of standard errors are close to the corresponding true standard errors for each approach across all the missing proportions (Tables 5.4, 5.5, 5.6, B.4, B.5).

The RGEE estimates are consistently more efficient than WGEE estimates for

Table 5.4: Simulation Setting 1
2x2 Crossover Study: comparison of the reweighted estimators with WGEE and other methods estimators under monotone missingness, where $20-30\%$ of the subjects were missing observations in period 2, and $\theta_1, \theta_2, \theta_3$ are intercept,treatment and baseline effect respectively, and the sample size is 1000.

	Bias			Sample SE			Mean Theoretical SE			95% CP		
Approach	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3
Large Treatment Effect (mis. pct., $20-30\%$)												
Full cohort	.002	.007	.005	.644	.286	.412	.645	.287	.416	.954	.955	.957
Complete-Case	-.134	-.128	.026	.739	.347	.499	.750	.357	.513	.918	.925	.938
WGEE(α)	.020	.029	.007	.830	.399	.439	.847	.413	.446	.943	.945	.952
WGEE($\hat{\alpha}$)	.019	.025	.007	.812	.396	.438	.828	.407	.444	.944	.948	.954
RGEE($\hat{\alpha}$)	.018	.023	.007	.717	.347	.431	.727	.353	.435	.943	.947	.952
Small Treatment Effect (mis. pct., $20-30\%$)												
Full cohort	.002	.007	.005	.648	.288	.415	.649	.289	.419	.954	.955	.957
Complete-Case	-.136	-.130	.026	.748	.351	.505	.759	.361	.519	.917	.924	.937
WGEE(α)	.020	.029	.007	.837	.402	.443	.854	.417	.450	.943	.945	.952
WGEE($\hat{\alpha}$)	.019	.025	.007	.818	.399	.441	.834	.410	.447	.944	.948	.954
RGEE($\hat{\alpha}$)	.018	.023	.007	.722	.349	.433	.732	.355	.438	.943	.947	.952

all missing proportions. Specifically, as shown in Table 5.4, the variance of RGEE improves as much as 25% over WGEE for the treatment effect in the case where $20-30\%$ of the subjects have missing observations in period 2. The improvement for the intercept is also significant, about 20%, whereas the improvement for baseline effect is modest. As the missing percentage increases to $40-50\%$ (Table 5.5) and $60-70\%$ (B.5), the magnitude of improvement increases to approximately 35% and 40% for the treatment effect respectively. It indicates that the proposed method is more stable than WGEE as the observation probability becomes small. As expected from the findings above, with missing proportion dropping to $10-15\%$ (B.4), the improvement decreases correspondingly. The influence of missing observations on baseline effect is quite limited. Not only is the bias small, but also the variations

Table 5.5: Simulation Setting 2
2x2 Crossover Study: comparison of the reweighted estimators with WGEE and other methods estimators under monotone missingness, where $40-50\%$ of the subjects were missing observations in period 2, and $\theta_1, \theta_2, \theta_3$ are intercept,treatment and baseline effect respectively, and the sample size is 1000.

	Bias			Sample SE			Mean Theoretical SE			95% CP		
Approach	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3
Large Treatment Effect (mis. pct., $40-50\%$)												
Full cohort	.005	.007	.005	.645	.287	.413	.646	.288	.417	.953	.954	.957
Complete-Case	-.733	-.250	.031	.923	.438	.524	1.01	.478	.563	.855	.915	.927
WGEE(α)	.031	.037	.009	1.08	.518	.461	1.21	.578	.495	.921	.935	.943
WGEE($\hat{\alpha}$)	.028	.034	.009	1.04	.508	.460	1.15	.545	.493	.935	.937	.945
RGEE($\hat{\alpha}$)	.022	.032	.009	.887	.409	.453	.949	.436	.476	.934	.935	.943
Small Treatment Effect (mis. pct., $40-50\%$)												
Full cohort	.005	.007	.005	.650	.289	.416	.651	.290	.420	.953	.954	.957
Complete-Case	-.747	-.255	.032	.937	.445	.532	1.02	.485	.571	.854	.914	.926
WGEE(α)	.032	.038	.009	1.09	.523	.465	1.22	.583	.500	.920	.934	.942
WGEE($\hat{\alpha}$)	.028	.035	.009	1.05	.512	.464	1.16	.549	.497	.935	.937	.945
RGEE($\hat{\alpha}$)	.022	.033	.009	.894	.412	.456	.956	.439	.479	.934	.935	.943

between sampling variance and sandwich estimator variances are negligible.

The magnitude of improvement in efficiency of RGEE relative to WGEE is rather remarkable for the treatment effect, which also happens to be the primary interest of the study. One of the possible explanations might be that unlike other covariates whose effect can be estimated from early periods, the treatment is distinctive in each period for a specific subject, and thus the missing information for the corresponding treatment would be difficult to replace and tends to be more damaging. As a result, it would be more difficult to estimate the treatment effect more precisely without proper adjustment. Our study also showed that the estimates from the data generated using small treatment coefficients showed trends similar to those from the large treatment effect.

Table 5.6: Simulation Setting 3
2x2 Crossover Study: comparison of the reweighted estimators with WGEE and other methods estimators under monotone missingness, where approximately $20-30\%$ of the subjects were missing observations in period 2, and $\theta_1, \theta_2, \theta_3$ are intercept,treatment and baseline effect respectively with large treatment effect and different sample size.

	Bias			Sample SE			Mean theoretical SE			95% CP		
Approach	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3
Cohort size 500												
Full Cohort	.002	.009	.005	.685	.303	.427	.691	.306	.431	.943	.945	.952
Complete-Case	-.149	-.131	.030	.852	.403	.512	.873	.411	.520	.806	.838	.840
WGEE(α)	.035	.031	.013	.955	.475	.455	.972	.483	.463	.852	.857	.873
WGEE($\hat{\alpha}$)	.033	.029	.011	.938	.469	.453	.955	.477	.462	.855	.861	.887
RGEE($\hat{\alpha}$)	.031	.027	.009	.839	.397	.444	.853	.403	.452	.861	.875	.893
Cohort size 1000												
Full cohort	.002	.007	.005	.644	.286	.412	.645	.287	.416	.954	.955	.957
Complete-Case	-.134	-.128	.026	.739	.347	.499	.750	.357	.513	.918	.925	.938
WGEE(α)	.020	.029	.007	.830	.399	.439	.847	.413	.446	.943	.945	.952
WGEE($\hat{\alpha}$)	.019	.025	.007	.812	.396	.438	.828	.407	.444	.944	.948	.954
RGEE($\hat{\alpha}$)	.018	.023	.007	.717	.347	.431	.727	.353	.435	.943	.947	.952

Compared to the estimators generated from data with 500 patients, those from the data with 1000 patients showed approximately $20-30\%$ improvement in variances for treatment effects for all the methods other than full cohort analysis, where the reduction is $10-15\%$, shown in Table 5.6. The large sample size case, the RGEE approach is more efficient than WGEE method even when only half of the patients were enrolled, where the improvements are about $20-30\%$ for intercept and the treatment effect.

Consistent with findings by Robins, Rotnitzky and Zhao (1995), we found that even when the nonresponse probabilities $\bar{\lambda}_{it}$ are known, the estimator $\hat{\theta}$, produced by WGEE($\hat{\alpha}$) that uses the estimated probabilities $\bar{\lambda}_{it}(\hat{\alpha})$ is at least as efficient as $\hat{\theta}^*$, the estimator produced by WGEE(α) that uses the true $\bar{\lambda}_{it}$.

3×3 Williams Design

For Williams design of 3×3 crossover study, we generated 500 replicated study data sets with sample sizes of $N = 900$ or 1800. As in the 2×2 crossover study, the period 1 outcome of each subject is always observed and the missingness is monotone. That is, period 3 is missing if period 2 is not observed. Covariates used in simulating the data include treatment, period, sex, and age and were never missing. The responses at period 2 and period 3 are observed at random according to the following logistic regression models:

$$\text{logit}\left\{P(R_{i2} = 1|X_i, Y_{i1})\right\} = \alpha_0 + \alpha_{1x}^T X_{i2} + \alpha_{1y} Y_{i1}, \tag{5.3}$$

$$\text{logit}\left\{P(R_{i3} = 1|R_{i2} = 1, X_i, Y_{i1}, Y_{i2})\right\} = \alpha_0 + \alpha_{2x}^T X_{i3} + \alpha_{1y} Y_{i1} + \alpha_{2y} Y_{i2}, \tag{5.4}$$

where X_{i2}, X_{i3} are covariates of subject i in period 2 and period 3, respectively, and Y_{i1}, Y_{i2} are the observed responses in period 1 and period 2, respectively.

We focused on data set with $40 - 50\%$ missing observations in the Williams 3x3 crossover study. Estimators generated by full cohort, CC, WGEE and RGEE are compared in terms of consistency, efficiency, and the coverage probability. Additionally, in order to evaluate the influence of the "working" covariance matrix, we compared the estimates by analyzing the data using compound symmetry and unstructured "working" variance-covariance respectively while the true variance-covariances between observations of period 1, period 2, and period 3 are compound symmetry. Also compared is the effect of sample size on the estimates. Specifically, performances of the estimators are evaluated for sample size of 900 or 1800.

The following matrix displays the compound symmetry covariance structure,

Table 5.7: Simulation Settings for 3×3 Crossover Design

Missing Prop.	Sample Size	Treatment Effect	Working Cov.	Table
$40-50\%$	900 *vs* 1800	Large	CS	5.8
$40-50\%$	900	Large	CS *vs* UN	5.9

which implies that the outcome of each time point or period assumes the same variance and the correlation between any two time points or periods is also the same.

$$Compound\ Symmetry\ (CS) : \sigma^2 \begin{pmatrix} 1 & & & \\ \rho & 1 & & \\ \vdots & \vdots & \ddots & \\ \rho & \rho & \cdots & 1 \end{pmatrix}$$

The more complicated unstructured variance-covariance structure is shown below, which specifies a completely general $T \times T$ covariance matrix parameterized directly in terms of variances and covariances.

$$Unstructured\ (UN) : \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1T} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1T} & \sigma_{2T} & \cdots & \sigma_T^2 \end{pmatrix}$$

The efficiency improvements of RGEE over WGEE vary depending the sample size and "working" covariance structure used. The variance reduction tends to be greater for the data of bigger sample size and a "working" covariance that is close to the true covariance, which is compound symmetry structure in this case. Specifically, the performance of RGEE is better than WGEE with $25-30\%$ improvement when

Table 5.8: Simulation Setting 4

Williams design of 3x3 crossover study: comparison of the reweighted estimators with WGEE and other methods estimators under monotone missingness, where overall approximately $40-50\%$ of the subjects were missing observations in period 2, period 3 or both, and $\theta_1, \theta_2, \theta_3, \theta_4$ are treatment 1, treatment 2, age, and sex effect respectively, the "working" covariance is compound symmetry (CS) structure while the true covariance structure is compound symmetry .

	Bias				Sample SE				Mean theoretical SE				95% CP			
Approach	θ_1	θ_2	θ_3	θ_4	θ_1	θ_2	θ_3	θ_4	θ_1	θ_2	θ_3	θ_4	θ_1	θ_2	θ_3	θ_4
Cohort size 900																
Full Cohort	.008	.002	-.048	.009	.351	.368	.391	.891	.385	.409	.431	.987	.960	.934	.940	.948
Complete-Case	-1.04	-.332	-.145	-1.87	.517	.530	.541	1.14	.569	.591	.591	1.26	.586	.832	.838	.685
WGEE($\hat{\alpha}$)	.117	.034	-.044	-.312	.491	.487	.509	1.11	.551	.543	.569	1.19	.786	.870	.900	0.901
RGEE($\hat{\alpha}$)	.035	-.010	-.023	-.191	.459	.457	.471	1.03	.506	.507	.536	1.11	.818	.905	.912	.908
Cohort size 1800																
Full Cohort	.006	.017	-.047	.009	.289	.306	.323	.743	.315	.335	.357	.813	.948	.938	.958	.952
Complete-Case	-1.02	-.260	-.121	-1.82	.435	.443	.452	.959	.478	.489	.497	1.07	.603	.903	.915	.752
WGEE($\hat{\alpha}$)	.090	.028	-.033	-.299	.413	.413	.429	.937	.449	.453	.471	1.03	.874	.906	.938	.902
RGEE($\hat{\alpha}$)	.014	-.009	-.017	-.172	.355	.353	.373	.819	.388	.391	.409	.897	.906	.914	.941	.918

Table 5.9: Simulation Setting 5

Williams design of 3x3 crossover study: comparison of the reweighted estimators with WGEE and other methods estimators under monotone missingness, where overall approximately 40% to 50% of the subjects were missing observations in period 2, period 3 or both, and $\theta_1, \theta_2, \theta_3, \theta_4$ are treatment 1, treatment 2, age, and sex effect respectively, and the sample size is 900 with true covariance structure as compound symmetry.

	Bias				Sample SE				Mean theoretical SE				95% CP			
Approach	θ_1	θ_2	θ_3	θ_4	θ_1	θ_2	θ_3	θ_4	θ_1	θ_2	θ_3	θ_4	θ_1	θ_2	θ_3	θ_4
Compound Symmetry Covariance (CS)																
Full Cohort	.008	.002	-.048	.009	.351	.368	.391	.891	.385	.409	.431	.987	.960	.934	.940	.948
Complete-Case	-1.04	-.332	-.145	-1.87	.517	.530	.541	1.14	.569	.591	.591	1.26	.586	.832	.838	.685
WGEE($\hat{\alpha}$)	.117	.034	-.044	-.312	.491	.487	.509	1.11	.551	.543	.569	1.19	.786	.870	.900	0.901
RGEE($\hat{\alpha}$)	.035	-.010	-.023	-.191	.459	.457	.471	1.03	.506	.507	.536	1.11	.818	.905	.912	.908
Unstructured Covariance (UN)																
Full Cohort	.008	.002	-.005	.010	.365	.385	.403	.928	.413	.443	.457	1.03	.960	.934	.940	.948
Complete-Case	-1.04	-.334	-.145	-1.87	.546	.557	.571	1.19	.613	.632	.641	1.35	.586	.832	.838	.686
WGEE($\hat{\alpha}$)	.160	.035	-.046	-.330	.519	.515	.539	1.17	.582	.587	.603	1.34	.776	.874	.904	.898
RGEE($\hat{\alpha}$)	.031	-.013	-.028	-.206	.497	.495	.512	1.12	.558	.563	.583	1.27	.836	.902	.919	.904

the sample size is 1800 and "working" covariance is compound symmetry (Table 5.8), followed by $10-15\%$ improvement for the study with 900 patients and compound symmetry covariance, and by $5-10\%$ improvement for the study with 900 patients and unstructured covariances (Table 5.9).

We observe that the efficiency improved approximately $30-40\%$ for all approaches when the sample size doubles. We also observe that the estimators are more efficient when the "working" covariances are closer to the true covariances of the data. Specifically, we observed that the estimators produced by the approaches using the compound symmetry covariance structure, which is the true covariance

of the simulated data, as "working" covariance are $10 - 15\%$ more efficient than the estimators that are obtained using the unstructured covariance as "working" covariance.

5.3 Application to Data

In this section, we illustrate the proposed method using the QTc study data. In cardiology, the QT interval is a measure of the time between the start of the Q wave and the end of the T wave in the heart's electrical cycle. In general, the QT interval represents electrical depolarization and repolarization of the left and right ventricles. QTc is the QT interval corrected for heart rates. A lengthened QT interval is a biomarker for ventricular tachyarrhythmias like torsades de pointes (TdP), which is a risk factor for sudden death. TdP was merely an esoteric diagnosis until the 1980s and 1990s when it was recognized as a major cause of drug-induced sudden cardiac death. Its recognition prompted the withdrawal of several popular medications from the market, including Sertindole, an antipsychotic drug, Cisapride, a GI prokinetic agent, Astemizole, a non-sedating antihistamine, and Grepafloxacin, an antibiotic. The regulatory guidance for pharmaceutical industry, ICH E14 clinical evaluation of QT/QTc interval prolongation and proarrhythmic potential for non-antiarrhythmic drugs, requires all sponsors submitting new drug applications to conduct a thorough QT/QTc study. A QT study often uses structured crossover designs such as Latin squares or Williams squares. When ignoring missing data, the structured design falls back into a random crossover design that suffers loss of

efficiency in addition to the sample size reduction. The replacement method, the commonly used method by pharmaceutical companies, replaces a missing subject with a newly recruited patient, but it is at odds with the idea of intention to treat (ITT). The data used here is a Williams design of 3×3 crossover study. It has three arms: placebo, tested drug, and the positive control, 400 mg Moxifloxacin, which is known to induce QT prolongation.

The primary model is:

$$E(Y_{it}|X_{it}) = \mu_{it} = \eta = X_{it}^T\theta, \tag{5.5}$$

where Y_{it} is the QTc outcome at period t, $t = 1, 2, 3$, and X_{it} are covariates. The MAR missing mechanism is assumed. We use a logistic model for the conditional probability $\bar{\lambda}_{it}$ at period t given that it is observed at period $t-1$. That is,

$$\bar{\lambda}_{it} = \text{logit}^{-1}(\alpha^T \bar{W}_{it}), \tag{5.6}$$

where $\bar{W}_{it}$ is the entire history of subject i before time t.

Complete-case, WGEE and RGEE methods are used to analyze the data, where about 36 patients are assigned to one of six treatment sequences with 6 patients in each treatment sequence. Approximately $20 - 30\%$ observations were missing. The coefficients estimated by fitting the model (5.6), shown in Table (5.10), are used to compute the observation probability. The estimates of the parameters in model (5.5) and their standard errors as calculated by the complete-case, WGEE, and RGEE methods are listed in Table 5.11.

Analysis of the QTc data with missing observations showed that the estimators of RGEE and WGEE methods were much more efficient than those produced by

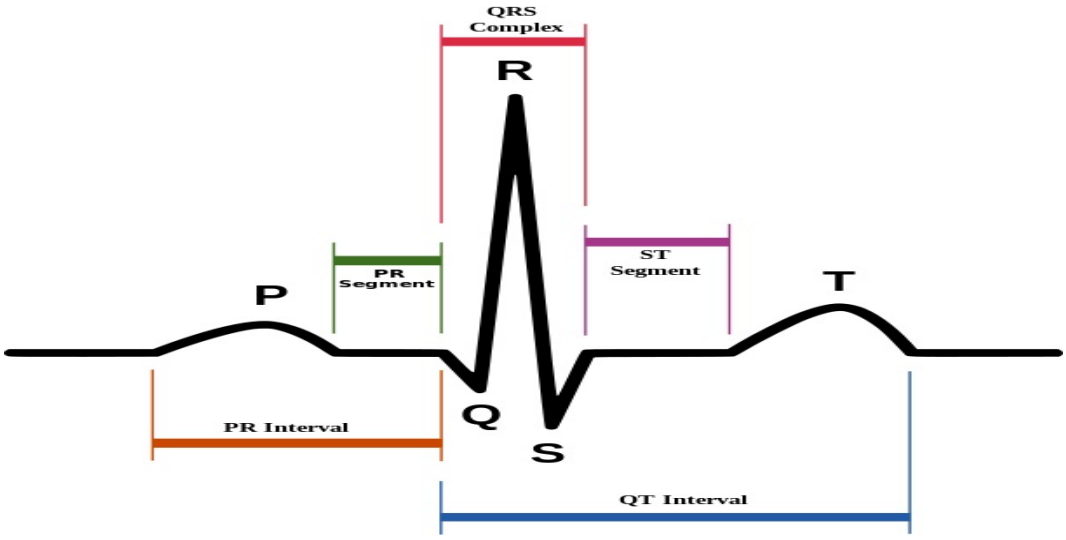

Figure 5.1: Schematic representation of Electrocardiogram (ECG)

Table 5.10: Estimated Logistic Model of the Observation Probability

	Period 2		Period 3	
Parameters	Estimate	SE	Estimate	SE
Treatment	-2.908	1.489	1.364	0.938
Age	-0.075	0.133	0.080	0.069
Sex (Male)	1.517	3.022	0.860	2.014
Race (White)	6.871	5.812	-2.021	2.383
Height	-0.942	1.071	-0.624	0.777
Weight	0.821	1.544	0.777	0.928
BMI	-0.656	5.277	-2.995	2.872
Y1 (outcome of Period 1)	0.535	0.964	0.319	0.804
Y2 (outcome of Period 2)	NA	NA	10.404	9.009

Table 5.11: Analysis of QTc Study Using Various Approaches

	Complete-case		WGEE		RGEE	
Effects	θ	SE	θ	SE	θ	SE
Drug	3.980	1.179	3.540	0.193	3.480	0.134
Moxi	11.95	3.534	10.390	0.711	8.640	0.658
Age	0.283	0.215	0.149	0.030	0.220	0.028
Sex (Male)	-1.21	3.376	-3.740	0.487	-3.990	0.478
Race (White)	0.051	3.381	-0.862	0.597	-1.213	0.558
BMI	0.211	0.588	0.595	0.107	0.560	0.101

complete-case analysis, and that RGEE consistently performed better than WGEE with the magnitude of improvement being greater for some effects. Specifically, for the drug coefficient estimate, the variance improvement of the RGEE method relative to the WGEE method is approximately 50%, and 14% for Moxi, the positive control.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

We have proposed the RGEE method, an approach that is more stable and efficient than WGEE. The estimates by the RGEE method have been shown consistent and asymptotically normal. We applied the RGEE method along with full cohort, complete-case, WGEE methods to simulated 2×2 crossover and 3×3 Williams design studies, and to real data. Both WGEE and RGEE estimates are consistent while complete-case estimates are biased, and the bias becomes more problematic as the missing proportion increases. Our study shows that the RGEE estimators are consistently more efficient than the WGEE estimators, and that the improvement is unequivocal and substantial although the magnitude of efficiency improvement differs as the sample size, missing proportion, or "working" covariance varies. Additionally, we find that for 2×2 crossover study, the estimates in the cohort size of 500 are fairly unbiased compared to those in the cohort size of 1000 while the estimates in the cohort size 900 of 3×3 Williams design are more biased than those in the cohort size of 1800. However the variance efficiency improves in both 2×2 and 3×3 simulations with the improvement being greater in 3×3 crossover study as sample size increases. Our study also shows that the efficiency improvement of RGEE relative to WGEE increases as the missing proportion becomes greater. Among

other findings is that the RGEE method performs better than the WGEE method regardless of the choice of "working" covariance. Choosing the "correct" covariance structure, however, does generally produce more efficient estimators. Consistent with findings by Robins, Rotnitzky and Zhao (1995), we found that even when the nonresponse probabilities $\bar{\lambda}_{it}$ are known, the estimator $\hat{\theta}$, produced by WGEE($\hat{\alpha}$) that uses the estimated probabilities $\bar{\lambda}_{it}(\hat{\alpha})$ is at least as efficient as $\hat{\theta}^*$, the estimator produced by WGEE(α) that uses the true $\bar{\lambda}_{it}$. Application to real data also showed that the RGEE method is more efficient than the WGEE method.

6.2 Future Work

Unlike longitudinal studies where a patient receives just one treatment and measured multiple times, crossover studies are more complicated because a patient receives a different treatment in each period. The reweighted generalized estimating equations method, although proposed for the crossover studies, can be applied to other longitudinal studies with slight modifications. As the methods of generalized estimating equations class have potential to handle the discrete variables well, it will be interesting to see how the proposed method performs when the missing outcome follows binary or other discrete distributions. Crossover studies where multiple measures are taken within each period are particularly difficult to deal with as the correlations also exist between the time points within each period in addition to the correlations between periods. It will be valuable to apply the proposed method to the studies of this nature. Further research regarding the performance

of the method under the missing not at random mechanism or the nonmonotone missingness pattern will be certainly informative.

Appendix A

Q-Q Plots and Histograms

A.1 Full Cohort Analysis

A.2 Complete-Case Analysis

A.3 WGEE Analysis

A.4 RGEE Analysis

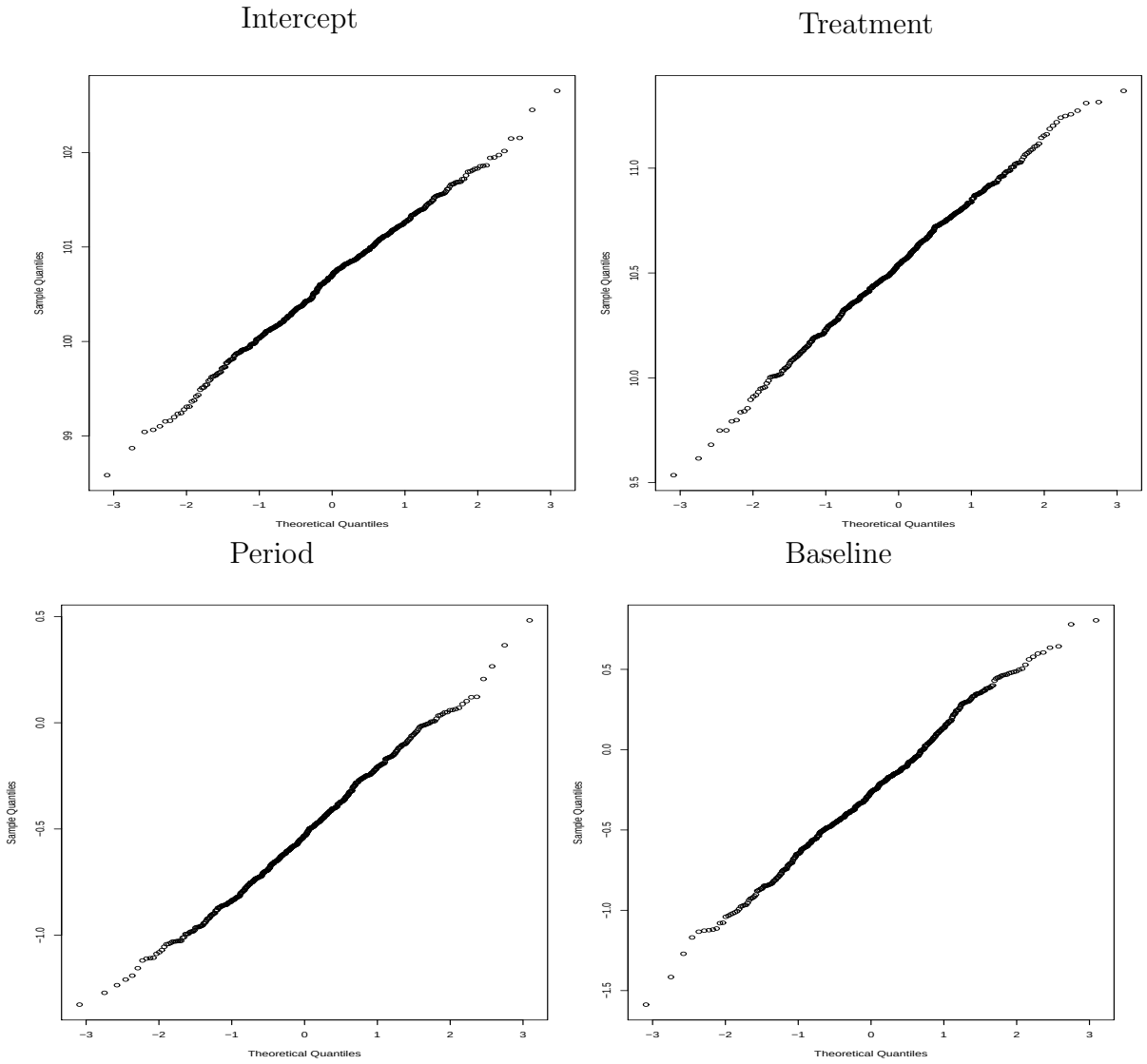

Figure A.1: Q-Q Plots of Full Cohort Analysis

2x2 crossover study, where $20 - 30\%$ of the subjects have missing observations in period 2 under monotone missingness. Coefficient estimates are for intercept, treatment, period, and baseline effects respectively. The sample size is 1000.

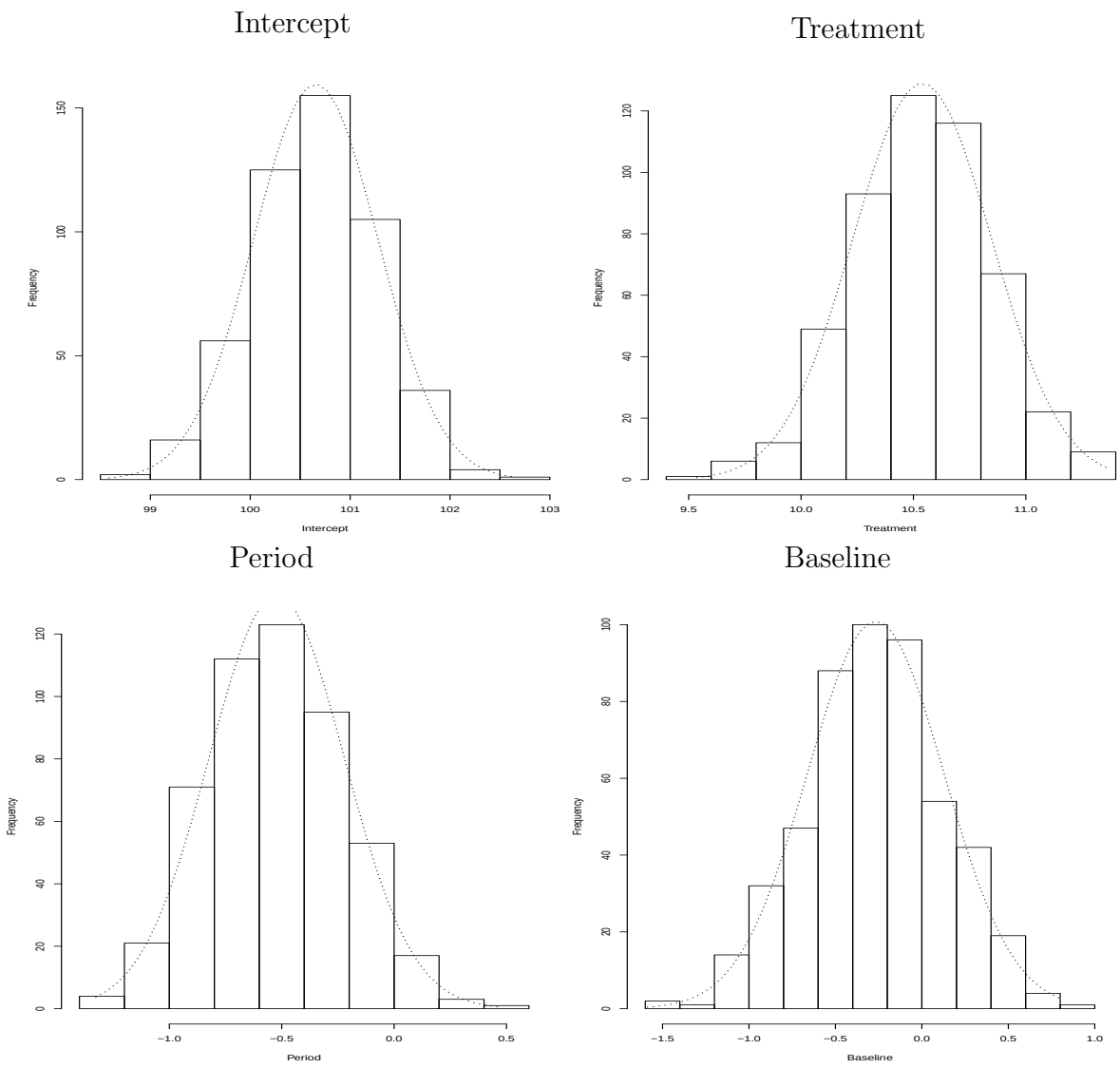

Figure A.2: Histograms of Full Cohort Analysis

2x2 crossover study, where $20 - 30\%$ of the subjects have missing observations in period 2 under monotone missingness. Coefficient estimates are for intercept, treatment, period, and baseline effects respectively. The sample size is 1000.

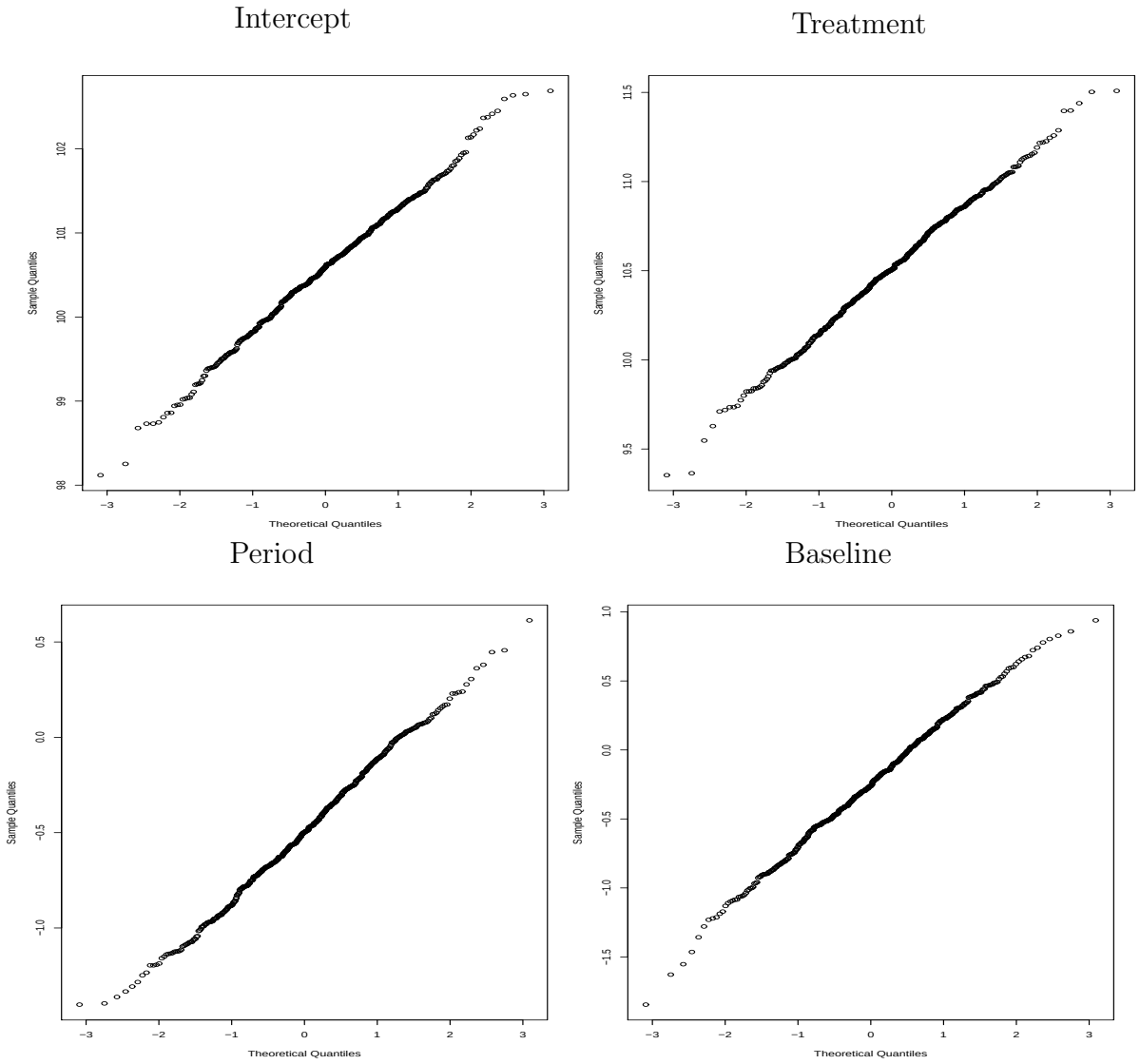

Figure A.3: Q-Q Plots of Complete-Case Analysis

2x2 crossover study, where $20 - 30\%$ of the subjects have missing observations in period 2 under monotone missingness. Coefficient estimates are for intercept, treatment, period, and baseline effects respectively. The sample size is 1000.

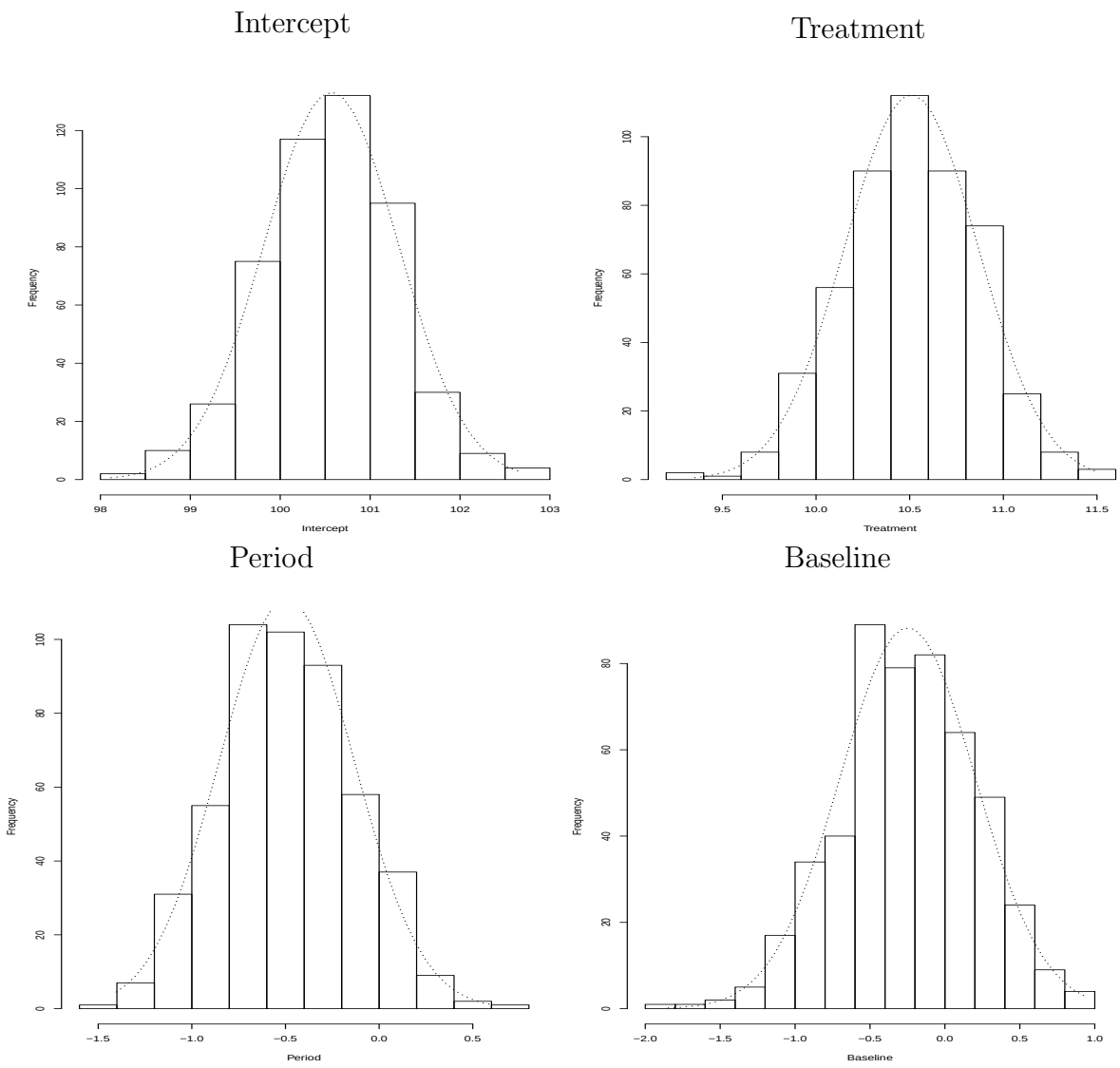

Figure A.4: Histograms of Complete-Case Analysis

2x2 crossover study, where $20 - 30\%$ of the subjects have missing observations in period 2 under monotone missingness. Coefficient estimates are for intercept, treatment, period, and baseline effects respectively. The sample size is 1000.

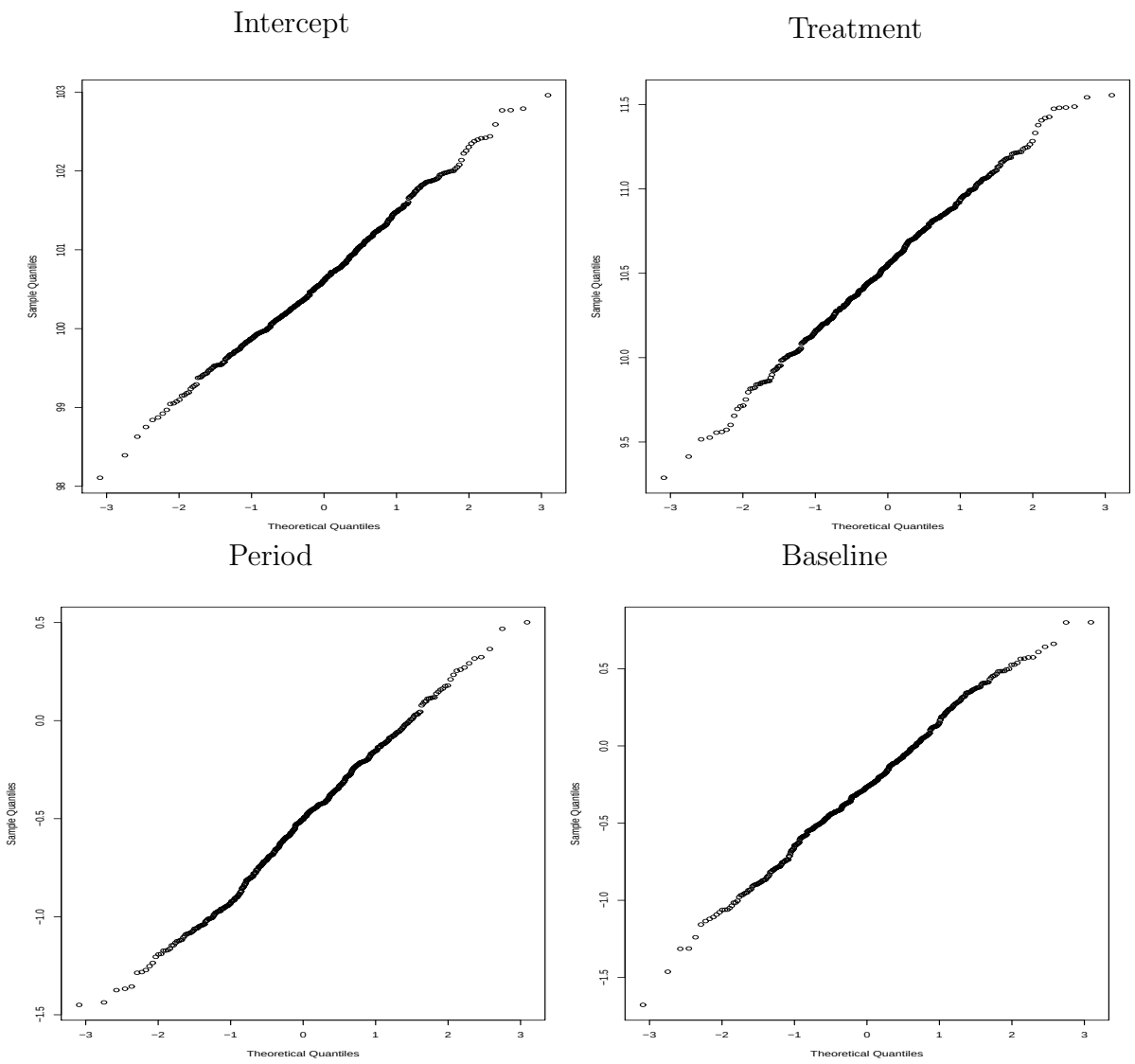

Figure A.5: Q-Q Plots of WGEE Analysis

2x2 crossover study, where $20 - 30\%$ of the subjects have missing observations in period 2 under monotone missingness. Coefficient estimates are for intercept, treatment, period, and baseline effects respectively. The sample size is 1000.

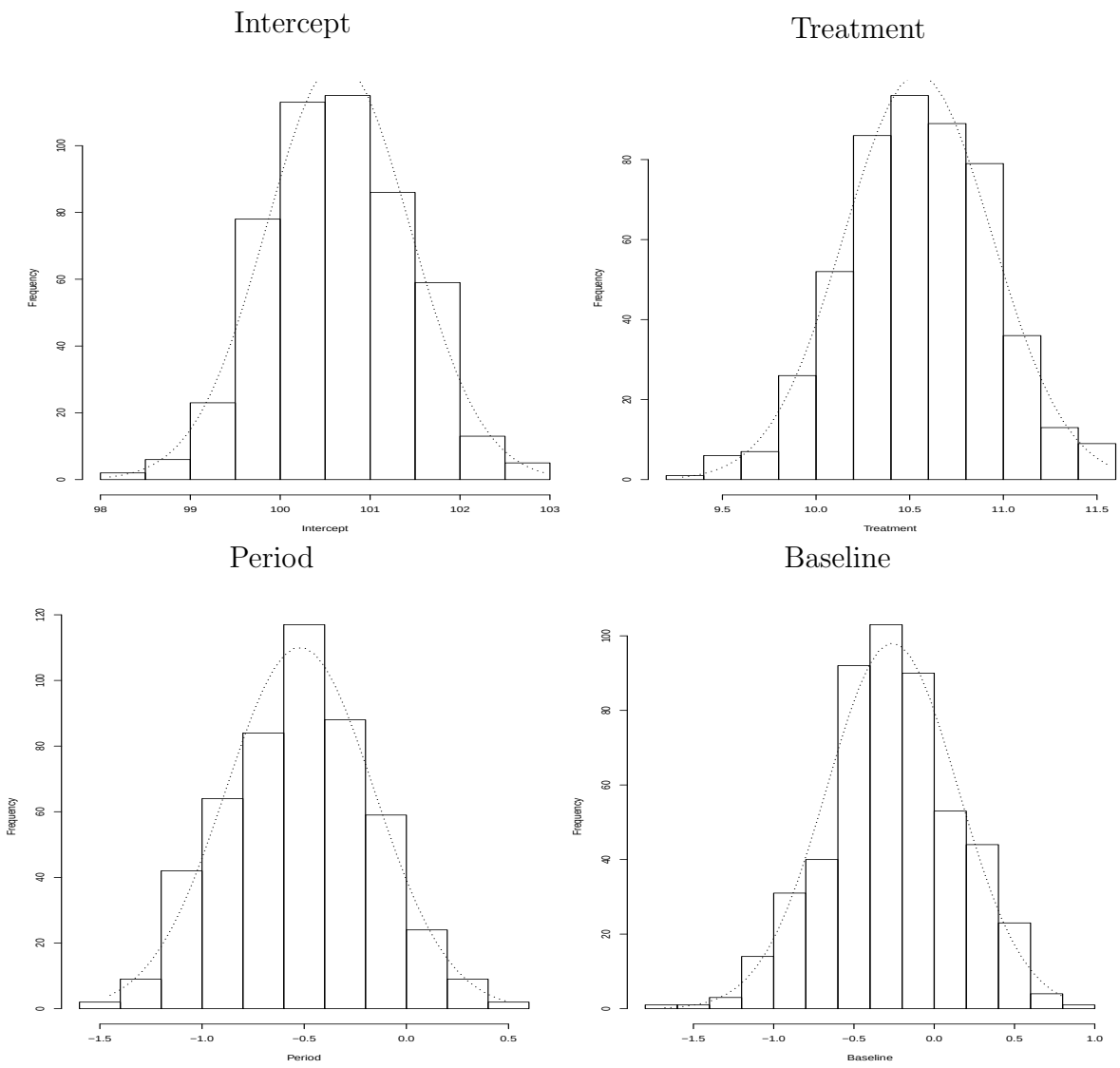

Figure A.6: Histograms of WGEE Analysis

2x2 crossover study, where $20-30\%$ of the subjects have missing observations in period 2 under monotone missingness. Coefficient estimates are for intercept, treatment, period, and baseline effects respectively. The sample size is 1000.

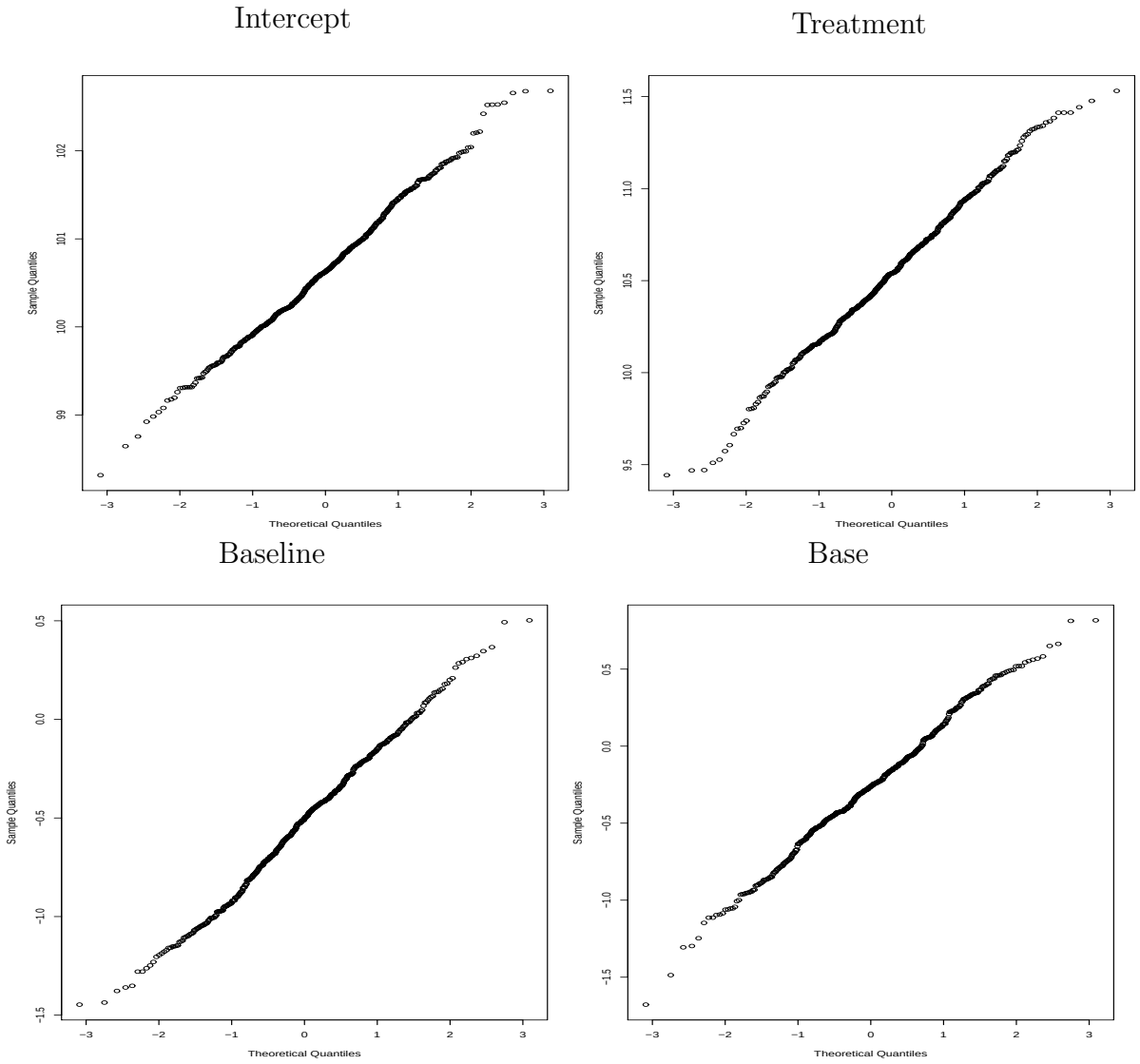

Figure A.7: Q-Q Plots of RGEE Analysis

2x2 crossover study, where $20 - 30\%$ of the subjects have missing observations in period 2 under monotone missingness. Coefficient estimates are for intercept, treatment, period, and baseline effects respectively. The sample size is 1000.

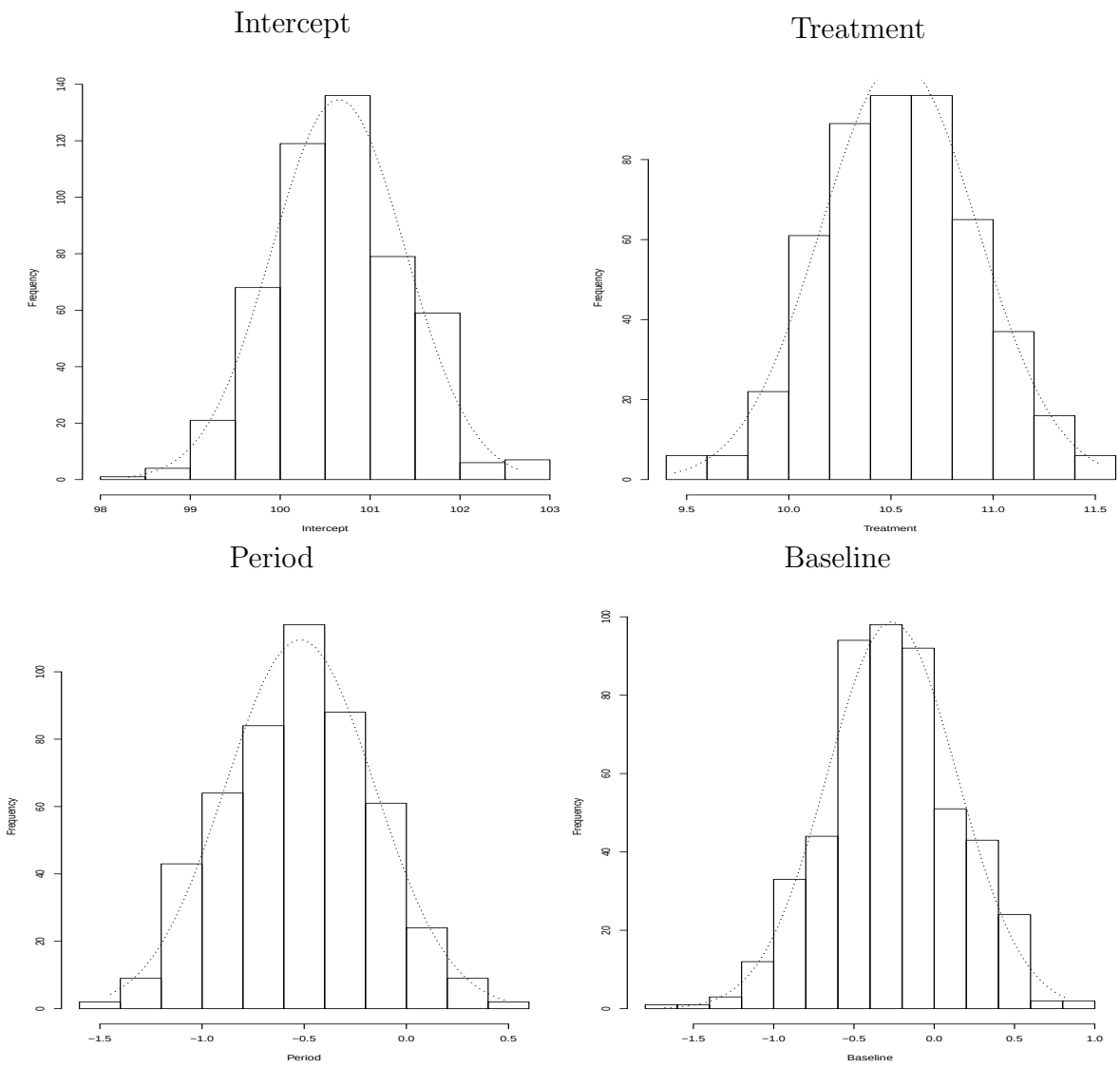

Figure A.8: Histograms of RGEE Analysis

2x2 crossover study, where $20-30\%$ of the subjects have missing observations in period 2 under monotone missingness. Coefficient estimates are for intercept, treatment, period, and baseline effects respectively. The sample size is 1000.

Appendix B

Simulation Results

B.1 2×2 *Crossover Design*

B.1.1 Simulation Setting 1

B.1.2 Simulation Setting 2

B.1.3 Simulation Setting 3

B.1.4 Simulation Setting 6

B.1.5 Simulation Setting 7

B.2 3×3 *Williams Design*

B.2.1 Simulation Setting 4

B.2.2 Simulation Setting 5

Table B.1: Simulation Setting 1

2x2 Crossover Study: comparison of the reweighting estimators with WGEE and other methods estimators under monotone missingness, where $20-30\%$ of the subjects were missing observations in period 2, and $\theta_1, \theta_2, \theta_3$ are intercept,treatment and baseline effect respectively, and the sample size is 1000.

	Bias			Sample SE			Mean Theoretical SE			95% CP		
Approach	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3
Large Treatment Effect (mis. pct., $20-30\%$)												
Full cohort	.002	.007	.005	.644	.286	.412	.645	.287	.416	.954	.955	.957
Complete-Case	-.134	-.128	.026	.739	.347	.499	.750	.357	.513	.918	.925	.938
WGEE(α)	.020	.029	.007	.830	.399	.439	.847	.413	.446	.943	.945	.952
WGEE($\hat{\alpha}$)	.019	.025	.007	.812	.396	.438	.828	.407	.444	.944	.948	.954
RGEE($\hat{\alpha}$)	.018	.023	.007	.717	.347	.431	.727	.353	.435	.943	.947	.952
Small Treatment Effect (mis. pct., $20-30\%$)												
Full cohort	.002	.007	.005	.648	.288	.415	.649	.289	.419	.954	.955	.957
Complete-Case	-.136	-.130	.026	.748	.351	.505	.759	.361	.519	.917	.924	.937
WGEE(α)	.020	.029	.007	.837	.402	.443	.854	.417	.450	.943	.945	.952
WGEE($\hat{\alpha}$)	.019	.025	.007	.818	.399	.441	.834	.410	.447	.944	.948	.954
RGEE($\hat{\alpha}$)	.018	.023	.007	.722	.349	.433	.732	.355	.438	.943	.947	.952

Table B.2: Simulation Setting 2

2x2 Crossover Study: comparison of the reweighting estimators with WGEE and other methods estimators under monotone missingness, where $40-50\%$ of the subjects were missing observations in period 2, and $\theta_1, \theta_2, \theta_3$ are intercept,treatment and baseline effect respectively, and the sample size is 1000.

	Bias			Sample SE			Mean Theoretical SE			95% CP		
Approach	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3
Large Treatment Effect (mis. pct., $40-50\%$)												
Full cohort	.005	.007	.005	.645	.287	.413	.646	.288	.417	.953	.954	.957
Complete-Case	-.733	-.250	.031	.923	.438	.524	1.01	.478	.563	.855	.915	.927
WGEE(α)	.031	.037	.009	1.08	.518	.461	1.21	.578	.495	.921	.935	.943
WGEE($\hat{\alpha}$)	.028	.034	.009	1.04	.508	.460	1.15	.545	.493	.935	.937	.945
RGEE($\hat{\alpha}$)	.022	.032	.009	.887	.409	.453	.949	.436	.476	.934	.935	.943
Small Treatment Effect (mis. pct., $40-50\%$)												
Full cohort	.005	.007	.005	.650	.289	.416	.651	.290	.420	.953	.954	.957
Complete-Case	-.747	-.255	.032	.937	.445	.532	1.02	.485	.571	.854	.914	.926
WGEE(α)	.032	.038	.009	1.09	.523	.465	1.22	.583	.500	.920	.934	.942
WGEE($\hat{\alpha}$)	.028	.035	.009	1.05	.512	.464	1.16	.549	.497	.935	.937	.945
RGEE($\hat{\alpha}$)	.022	.033	.009	.894	.412	.456	.956	.439	.479	.934	.935	.943

Table B.3: Simulation Setting 3

2x2 Crossover Study: comparison of the reweighting estimators with WGEE and other methods estimators under monotone missingness, where approximately $20-30\%$ of the subjects were missing observations in period 2, and $\theta_1, \theta_2, \theta_3$ are intercept,treatment and baseline effect respectively with different sample size and large treatment effect size.

	Bias			Sample SE			Mean theoretical SE			95% CP		
Approach	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3
Cohort size 500												
Full Cohort	.002	.009	.005	.685	.303	.427	.691	.306	.431	.943	.945	.952
Complete-Case	-.149	-.131	.030	.852	.403	.512	.873	.411	.520	.806	.838	.840
WGEE(α)	.035	.031	.013	.955	.475	.455	.972	.483	.463	.852	.857	.873
WGEE($\hat{\alpha}$)	.033	.029	.011	.938	.469	.453	.955	.477	.462	.855	.861	.887
RGEE($\hat{\alpha}$)	.031	.027	.009	.839	.397	.444	.853	.403	.452	.861	.875	.893
Cohort size 1000												
Full cohort	.002	.007	.005	.644	.286	.412	.645	.287	.416	.954	.955	.957
Complete-Case	-.134	-.128	.026	.739	.347	.499	.750	.357	.513	.918	.925	.938
WGEE(α)	.020	.029	.007	.830	.399	.439	.847	.413	.446	.943	.945	.952
WGEE($\hat{\alpha}$)	.019	.025	.007	.812	.396	.438	.828	.407	.444	.944	.948	.954
RGEE($\hat{\alpha}$)	.018	.023	.007	.717	.347	.431	.727	.353	.435	.943	.947	.952

Table B.4: Simulation Setting 6

2x2 Crossover Study: comparison of the reweighted estimators with WGEE and other methods estimators under monotone missingness, where $10-15\%$ of the subjects were missing observations in period 2, and $\theta_1, \theta_2, \theta_3$ are intercept,treatment and baseline effect respectively, and the sample size is 1000.

	Bias			Sample SE			Mean Theoretical SE			95% CP		
Approach	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3
Large Treatment Effect (mis. pct., $10-15\%$)												
Full cohort	.002	.007	.005	.640	.285	.410	.641	.286	.414	.954	.956	.958
Complete-Case	-.081	-.017	.012	.662	.313	.494	.673	.319	.499	.939	.941	.948
WGEE(α)	.010	.010	.005	.726	.370	.431	.739	.385	.443	.948	.951	.952
WGEE($\hat{\alpha}$)	.009	.011	.006	.712	.369	.430	.725	.373	.438	.949	.952	.955
RGEE($\hat{\alpha}$)	.008	.012	.005	.678	.334	.419	.691	.341	.435	.948	.951	.954
Small Treatment Effect (mis. pct., $10-15\%$)												
Full cohort	.002	.007	.005	.644	.286	.412	.645	.287	.416	.954	.956	.958
Complete-Case	-.082	-.017	.012	.669	.316	.499	.680	.322	.504	.939	.941	.948
WGEE(α)	.010	.010	.005	.732	.373	.434	.745	.388	.447	.948	.951	.952
WGEE($\hat{\alpha}$)	.009	.011	.006	.717	.372	.433	.730	.376	.441	.949	.952	.955
RGEE($\hat{\alpha}$)	.008	.012	.005	.682	.336	.422	.695	.343	.438	.948	.951	.954

Table B.5: Simulation Setting 7

2x2 Crossover Study: comparison of the reweighted estimators with WGEE and other methods estimators under monotone missingness, where $60-70\%$ of the subjects were missing observations in period 2, and $\theta_1, \theta_2, \theta_3$ are intercept,treatment and baseline effect respectively, and the sample size is 1000.

	Bias			Sample SE			Mean Theoretical SE			95% CP		
Approach	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3
Large Treatment Effect (mis. pct., $60-70\%$)												
Full cohort	.005	.007	.005	.650	.289	.416	.651	.290	.420	.952	.954	.955
Complete-Case	-2.12	-.825	-.015	1.16	.617	.78	1.23	.621	.793	.575	.720	.828
WGEE(α)	.071	.052	.017	1.57	.798	.521	1.78	.868	.512	.917	.925	.933
WGEE($\hat{\alpha}$)	.065	.047	.014	1.46	.779	.505	1.65	.848	.505	.927	.937	.942
RGEE($\hat{\alpha}$)	.062	.045	.011	1.23	.609	.491	1.32	.661	.485	.923	.938	.945
Small Treatment Effect (mis. pct., $60-70\%$)												
Full cohort	.005	.007	.005	.653	.291	.418	.654	.292	.423	.952	.954	.955
Complete-Case	-2.16	-.840	-.015	1.18	.627	.792	1.25	.631	.806	.574	.719	.827
WGEE(α)	.072	.053	.017	1.58	.805	.526	1.79	.876	.517	.916	.924	.932
WGEE($\hat{\alpha}$)	.066	.048	.014	1.47	.785	.509	1.66	.855	.509	.926	.936	.941
RGEE($\hat{\alpha}$)	.063	.046	.011	1.23	.613	.495	1.33	.666	.489	.923	.938	.945

Table B.6: Simulation Setting 4

Williams design of 3x3 crossover study: comparison of the reweighted estimators with WGEE and other methods estimators under monotone missingness, where overall approximately $40-50\%$ of the subjects were missing observations in period 2, period 3 or both, and $\theta_1, \theta_2, \theta_3, \theta_4$ are treatment 1, treatment 2, age, and sex effect respectively, the "working" covariance is compound symmetry (CS) structure while the true covariance structure is compound symmetry .

	Bias				Sample SE				Mean theoretical SE				95% CP			
Approach	θ_1	θ_2	θ_3	θ_4	θ_1	θ_2	θ_3	θ_4	θ_1	θ_2	θ_3	θ_4	θ_1	θ_2	θ_3	θ_4
Cohort size 900																
Full Cohort	.008	.002	-.048	.009	.351	.368	.391	.891	.385	.409	.431	.987	.960	.934	.940	.948
Complete-Case	-1.04	-.332	-.145	-1.87	.517	.530	.541	1.14	.569	.591	.591	1.26	.586	.832	.838	.685
WGEE($\hat{\alpha}$)	.117	.034	-.044	-.312	.491	.487	.509	1.11	.551	.543	.569	1.19	.786	.870	.900	0.901
RGEE($\hat{\alpha}$)	.035	-.010	-.023	-.191	.459	.457	.471	1.03	.506	.507	.536	1.11	.818	.905	.912	.908
Cohort size 1800																
Full Cohort	.006	.017	-.047	.009	.289	.306	.323	.743	.315	.335	.357	.813	.948	.938	.958	.952
Complete-Case	-1.02	-.260	-.121	-1.82	.435	.443	.452	.959	.478	.489	.497	1.07	.603	.903	.915	.752
WGEE($\hat{\alpha}$)	.090	.028	-.033	-.299	.413	.413	.429	.937	.449	.453	.471	1.03	.874	.906	.938	.902
RGEE($\hat{\alpha}$)	.014	-.009	-.017	-.172	.355	.353	.373	.819	.388	.391	.409	.897	.906	.914	.941	.918

Table B.7: Simulation Setting 5

Williams design of 3x3 crossover study: comparison of the reweighted estimators with WGEE and other methods estimators under monotone missingness, where overall approximately 40% to 50% of the subjects were missing observations in period 2, period 3 or both, and $\theta_1, \theta_2, \theta_3, \theta_4$ are treatment 1, treatment 2, age, and sex effect respectively, and the sample size is 900 with true covariance structure as compound symmetry.

	Bias				Sample SE				Mean theoretical SE				95% CP			
Approach	θ_1	θ_2	θ_3	θ_4	θ_1	θ_2	θ_3	θ_4	θ_1	θ_2	θ_3	θ_4	θ_1	θ_2	θ_3	θ_4
Compound Symmetry Covariance (CS)																
Full Cohort	.008	.002	-.048	.009	.351	.368	.391	.891	.385	.409	.431	.987	.960	.934	.940	.948
Complete-Case	-1.04	-.332	-.145	-1.87	.517	.530	.541	1.14	.569	.591	.591	1.26	.586	.832	.838	.685
WGEE($\hat{\alpha}$)	.117	.034	-.044	-.312	.491	.487	.509	1.11	.551	.543	.569	1.19	.786	.870	.900	0.901
RGEE($\hat{\alpha}$)	.035	-.010	-.023	-.191	.459	.457	.471	1.03	.506	.507	.536	1.11	.818	.905	.912	.908
Unstructured Covariance (UN)																
Full Cohort	.008	.002	-.005	.010	.365	.385	.403	.928	.413	.443	.457	1.03	.960	.934	.940	.948
Complete-Case	-1.04	-.334	-.145	-1.87	.546	.557	.571	1.19	.613	.632	.641	1.35	.586	.832	.838	.686
WGEE($\hat{\alpha}$)	.160	.035	-.046	-.330	.519	.515	.539	1.17	.582	.587	.603	1.34	.776	.874	.904	.898
RGEE($\hat{\alpha}$)	.031	-.013	-.028	-.206	.497	.495	.512	1.12	.558	.563	.583	1.27	.836	.902	.919	.904

Bibliography

[1] Brick, J.M. and Kalton, G. (1996). Handling Missing Data in Survey Research. *Stat. Meth. Med. Res.*, 5, 215-238.

[2] Carpenter, J.R., Kenward, M.G., and Vansteelandt, S. (2006). A Comparison of Multiple Imputation and Doubly Robust Estimation for Analyses with Missing Data. *J.R. Stat. Soc. Series A*, 169, 571-584.

[3] Horvitz. D. G. and Thompson, D. J. (1952). A Generalization of Sampling Without Replacement From a Finite Universe. *J Am. Stat. Assoc.*, 47, 663-685.

[4] Jones, B. and Kenward, M.G. (2003). *Design and Analysis of Cross-over Trials*, Chapman & Hall/CRC.

[5] Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation*, Second Edition, Springer.

[6] Liang, K.Y. and Zeger, S.L. (1986). Longitudinal Data Analysis using Generalized Models. *Biometrika*, 73, 13-22.

[7] Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*, Chichester: Wiley.

[8] McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*, Second Edition, Chapman & Hall/CRC.

[9] Molenberghs, G. and Kenward, M.G. (2007). *Missing Data in Clinical Studies*, Chichester, Wiley.

[10] Newey, W.K. and McFadden, D. (1994). Large Samples Estimation and Hypothesis Testing. *Handbook of Econometrics*, Vol. 4, eds. D. McFadden and R. Engler, Elsevier Science.

[11] National Research Council (2010). *The Prevention and Treatment of Missing Data in Clinical Trials*, Panel on Handling Missing Data in Clinical Trials. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academic Press.

[12] Pierce, D.A. (1982). The Asymptotic Effect of Subsituting Estimators for Parameters in Certain Types of Statistics. *The Annals of Statistics*, 10, 475-478.

[13] Prentice, R.L. (1988). Correlated Binary Regeression with Covariates Specific to each Binary Observation. *Biometrics*, 44, 1033-1048.

[14] Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of Regression Coefficients When Some Regressors Are Not Always Observed. *J Am. Stat. Assoc.*, 89, 846-866.

[15] Robins, J. M., Rotnizky, A., and Zhao, L.P. (1995). Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data. *J Am. Stat. Assoc.*, 90, 106-121.

[16] Rubin, D. B. (1976). Inference and Missing Data. *Biometrika*, 63, 581-592.

[17] Rubin, D.B. (1978). Multiple Imputations in Sample Surveys. *Proc. Survey Res. Meth. Sec., Am. Statist. Assoc*, 20-34.

[18] Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*, London: Chapman and Hall.

[19] Schafer, J. L. (1999). Multiple Imputation: a primer. *Stat. Meth. Med. Res.*, 8, 3-15.

[20] Wang, C.Y., Wang, S., Zhao, L.P., Qu, S.T. (1997). Wighted Semiparametric estimattion in regression analysis with missing covariate data. *J. Am. Stat. Assoc.*, 92, 512-525.

[21] Williams, E.J. (1949). Experimental Designs Balanced for the Estimation of Residual Effect of Treatments. *Aus. J. Sci. Res.*, 2, 149-168.

[22] Zhao, L. P., Lipsitz, S (1992). Designs and Analysis of Two-Stage Studies. *Statistics in Medicine*, 11, 769-782.